Individual Differences in Linguistic Prediction in Native Language Comprehension and Second Language Learning

Margarita Zeitlin

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2020

Reading Committee:
Lee Osterhout, Chair
Chantel Prat
Ione Fine

Program Authorized to Offer Degree:
Psychology

University of Washington

Abstract

Individual Differences in Linguistic Prediction in Native Language Comprehension and Second Language Learning

Margarita Zeitlin

Chair of the Supervisory Committee:

Lee Osterhout

Department of Psychology

Individual differences ERP research shows that grammatical errors violate expectations of form (P600 effects) in some individuals and expectations of meaning (N400 effects) in others. These differences have not been fully accounted for by variability in working memory capacity or language experience. Two experiments were conducted to investigate whether the observed variability in grammar processing can be explained by differences in comprehenders' grammar aptitude or in the way individuals build predictions during language comprehension. In Experiment 1, EEG data was recorded while native (L1) English speakers read sentences containing syntactic violations of various types, and semantic continuations that varied in expectancy. In Experiment 2, EEG data was recorded while native Chinese speakers read Chinese sentences containing one type of syntactic violation, and English sentences containing one syntactic violation familiar to the L1 and one violation unique to the second language (L2). The EEG signal was analyzed for differences in ERP response magnitudes and changes in beta power desynchronization in response to violations. English speakers’ ERP response patterns to syntactic anomalies predicted neural sensitivity to semantic anomalies. Chinese learners’ proficiency in the unique L2 rule was predicted by their grammar processing patterns in Chinese. Grammar processing variability was not explained by grammar aptitude, working memory, or language experience scores in either experiment. ERP results and preliminary data from EEG analyses from both experiments provide converging evidence that individuals may build stronger semantically or syntactically driven top-down linguistic predictions when supported by contextual constraints. Given that better predictions can lead to more efficient language processing and better language learning, these data reveal a feature of individuals’ intrinsic behavior that may be leveraged to improve spoken and written language comprehension and acquisition.

Acknowledgements

I have been very fortunate to receive a great deal of support throughout the development, execution, and writing of this dissertation, which was made possible in part by funding from the American Psychological Association Dissertation Research Award and the Department of Psychology.

I would like to extend my sincere thanks to Dr. Lee Osterhout who welcomed me to the University of Washington and gave me the opportunity to make Seattle my home. You gifted me the space to grow into my own as an independent researcher and motivated me to cultivate collaborative relationships that inspired the core of my dissertation work. Thank you also to Dr. Judith McLaughlin for your insight, kindness, and encouragement as I began to put down my academic roots. Your resourcefulness and commitment to our science provided a foundation that allowed me to flourish in new directions.

I am deeply indebted to Dr. Chantel Prat whose endless enthusiasm, inspiration, and confidence in my strengths and potential made every challenge an opportunity for growth. Any student who is lucky enough to be nurtured and molded by your hands understands what a rare and wonderful thing your attention, advice, and camaraderie is. I extend my most heartfelt appreciation to Dr. Andrea Stocco and the rest of the CCDL family for complementing the rigor of our scientific endeavors with the joy and delight of your friendship. You all brightened these last few years, enough to light the end of the tunnel and help me move toward it.

A special thanks goes out to Patrick Rice, without whom the completion of the time frequency analyses would not have been possible. I would also like to acknowledge the assistance of Dr. Mark Pettet for ensuring the smooth operation of the data collection lab and for your thought, care, and help in developing data processing tools to meet my study needs. I am especially grateful for my fellow graduate student Iris Kuo and my research assistants for the effort and industriousness you applied toward writing such perfect stimuli and the 750+ hours (!) of data collection that made the bulk of this dissertation possible. Thank you, Elaine Currie, Kathryn Duffy-Greslo, Alex Feng, Wesley Ganz, Yanfei Han, Cindy Hu, Vanessa Lin, Wenjun Ma, Naomi MacFarland, Tyler Muromoto, Justin Shan, and Chen Su.

I would also like to extend my gratitude to my committee members Dr. Ione Fine and Dr. Amy Pace for your perspective and engagement throughout the dissertation process. I would especially like to thank Dr. Akira Omaki for the immersive discussions we had in the early days

of my thesis development. In the short time we worked together, your intuition and curiosity sharpened my thinking and challenged me to reach beyond my assumptions to do the best science I could do. Your questions helped formulate the seeds of my dissertation work, and I would like to think that you would be just as intrigued today by the fruits they bore. You are deeply missed.

My deepest appreciation goes out to my parents, Stas and Inga, for their unparalleled support and for putting it into my head early on that I was capable of big things. Your invaluable contributions and guidance were instrumental in my success, and I thank you for teaching me to be resilient, diligent, and committed in all that I do. Thank you to my sister, Michelle, whose laughter and compassion lifted me up and gave me the resolve to push through with grace and a lighthearted spirit. My heartfelt thanks go out to my family for their profound belief in my abilities and their unwavering love.

Finally, my success would not have been possible without the nurturing and support of my future husband, Nicholas Stejer. This journey brought us together the day it began, and you have been by my side through it every day since. It was fitting that you should be the first pilot subject of my dissertation, because you are my co-pilot every day. Thank you for lifting me up by always reminding me of all I am capable of, and for grounding me when I need help centering on the important things. This dissertation is as much a product of your love and devotion as it is of my efforts, and I thank you.

Table of Contents

List of Figures

List of Tables

Chapter 1. INTRODUCTION

Language comprehension is no trivial pursuit. In order to successfully extract meaning from language, the brain must wrangle a noisy, ongoing signal that arrives incrementally at a brisk rate of 2-3 words per second. The creativity of language further impedes processing speed because the content of the signal is often difficult to anticipate, and multiple levels of linguistic analysis are necessary for comprehension. The fact that humans manage to understand each other despite this challenge is remarkable.

Part of the brain's solution to this problem lies in using the context of the ongoing message to build meaning- and structure-related predictions about what it might encounter next. The better its predictions, the more efficient its comprehension. However, not all comprehenders are made equal. Because differences in language skills have been shown to manifest in varied degrees of academic success and employment outcomes (*Early Skills and Predictors of Academic Success*, 2016; *Predictors of Postsecondary Success*, 2013; *The State of Languages in the U.S.: A Statistical Portrait*, 2016), the mechanisms behind linguistic predictions are of considerable societal importance. Much of what we learn throughout our lifetimes depends on written or spoken comprehension skills, especially when comprehension takes place in a non-native language. Understanding how individuals build successful linguistic predictions can thus help inform interventions for improving language skills in both native (L1) speakers and second language (L2) learners. This dissertation investigates individual variability in the type and strength of predictions made during L1 or L2 language processing, with the goal that such interventions could potentially leverage individuals' intrinsic processing strengths to promote better language outcomes.

1.1 ERP measures of language processing

Event-related potential (ERP) research on language comprehension has established that the average brain responds in qualitatively different ways to violations of semantic (i.e., meaning-based) and syntactic (i.e., form-based) expectations. ERPs are derived from recordings of electrical activity made from the surface of the scalp using electroencephalography (EEG). The millisecond-by-millisecond resolution captures fluctuations in neural activity as participants read sentences and reflects the ongoing temporal dynamics of language processing. For example, every word a reader encounters elicits an N400 ERP component, a centroparietally distributed waveform that follows the onset of that word and peaks in the negative direction around 400 ms. The N400 is sensitive to the lexical properties of words; words that appear with greater frequency in spoken or written corpora words elicit smaller N400 magnitudes than less frequent words (Rugg, 1990; Van Petten & Kutas, 1990), as do words belonging to certain grammatical categories (Münte et al., 2001; Vigliocco et al., 2011). Pertinent to the present study, there is also robust evidence that N400 magnitude is sensitive to the semantic fit of a word in its context. The N400 effect increases in magnitude as the likelihood of a sentence continuation for a given context, or cloze probability, decreases, such that highly expected words produce smaller N400 magnitudes than unexpected continuations, which in turn produce smaller N400 magnitudes than improbable semantic continuations, regardless of how constraining the context is (Kuperberg et al., 2020; Kutas & Federmeier, 2011; Kutas & Hillyard, 1984). Traditionally, the N400 effect has thus been interpreted as reflecting processes related to lexical access and semantic analysis and integration (Berkum et al., 1999; Borovsky et al., 2012; Lau et al., 2008; West & Holcomb, 2002).

The physical characteristics (i.e., latency, distribution, and direction) and functional role of the N400 contrast with those of the P600 ERP component. The P600 is a posteriorly distributed positive-going deflection in the waveform that begins to peak around 600 ms following word onset. The magnitude of the P600 increases in response to a variety of ambiguous or complex syntactic structures (i.e., garden path sentences, object-relative clauses, etc.) (Gouvea et al., 2010; Kaan & Swaab, 2003; Osterhout, 1997; Osterhout et al., 1994, 1994; Osterhout & Holcomb, 1993) or syntactic and morpho-syntactic violations (i.e., subject-verb agreement errors, phrase structure violations, etc.) (Friederici et al., 1996; Hagoort et al., 1993, 2003; Mehravari et al., 2015; Osterhout, 1997; Osterhout & Holcomb, 1992; Osterhout & Mobley, 1995). Though there is less consensus as to the exact function of the P600, it has typically been understood to reflect active syntactic processing related to integration, reanalysis or repair of sentence structure as a result of syntactic anomaly, ambiguity, or complexity (Friederici, 2002; Friederici et al., 2002; Kaan et al., 2000; Kaan & Swaab, 2003; Kos et al., 2010; Osterhout & Nicol, 1999).

1.2 The P600 as a marker of proficient grammar comprehension

The distinction between the N400 effect and the P600 effect established by group-level analyses has long served as a useful framework for investigating when and to what degree semantic or syntactic processing mechanisms are recruited for comprehension. The P600 effect, in particular, has come to be understood as the signature of a matured, efficient, and functioning grammar comprehension system in healthy, adult native speakers. For example, increased P600 magnitudes in native speakers have been related to improvements in grammar processing (Canette et al., 2020) and greater native language proficiency (Pakulak & Neville, 2010). Furthermore, studies of grammar processing in children (Clahsen et al., 2007; Hahne et al., 2004), individuals with Selective Language Impairment (Friederici, 2006; Purdy et al., 2014), and patients with aphasia (Kielar et al., 2012; Wassenaar et al., 2004) or Parkinson's disease (Friederici et al., 2003) provide evidence that an underdeveloped or impaired grammar comprehension system manifests in an absence, attenuation, or delay of the P600 effect to syntactic anomalies compared to healthy, adult controls.

Nowhere is the importance of the P600 as a marker of successful grammar processing more evident than in the realm of L2 research. The absence or attenuation of the P600 effect is generally taken as an indication that L2 learners have not yet achieved native-like grammar proficiency. Language learners at lower levels of proficiency typically show smaller or absent P600 effects to syntactic violations than learners with high L2 proficiency (Kotz, 2009; Rossi et al., 2006a; Tanner et al., 2013; van Hell & Tokowicz, 2010; White et al., 2012). Moreover, anomalies that involve L2 grammatical features that are not realized in the learner's L1 sometimes do not elicit robust P600 effects like they do in native speakers of the L2 (Chen et al., 2007; Ojima et al., 2005; Sabourin & Stowe, 2008; but see Steinhauer, 2014). These findings have been used to suggest that a matured language system is unable to effectively accommodate unfamiliar grammatical structures.

In addition to the magnitude of the P600 effect increasing in amplitude with increases in language proficiency, there is also evidence that children and adult learners' neural responses change qualitatively as a learner becomes increasingly proficient in the language (Clahsen et al., 2007; Hahne et al., 2004; McLaughlin et al., 2010; Osterhout et al., 2006, 2008; Tanner et al., 2013). McLaughlin and colleagues (2010) examined ERP responses to subject-verb agreement errors (e.g., *The books <u>are</u>/book <u>*are</u> heavy*)[1] in L2 French learners over their first year of

---

[1] An asterisk (*) is linguistic notation of ungrammaticality. When an asterisk appears before a word, it signifies that the word is considered ungrammatical given the context. Underlined words indicate the critical word that data is time-locked to in order to assess changes in brain activity following the critical word.

instruction. Four weeks into instruction, learners' ERPs were characterized by N400 responses, but by 24 weeks, L2 learners showed the expected P600 effect. According to the authors, this transition reflected the process by which new grammatical rules become instantiated in the brain. Learners initially rely on instance-based learning, so L2 grammars are somewhat idiomatic; grammatical violations appear more like phrases using the wrong words rather than incorrectly formed rules, thus eliciting an N400 effect. Over time, learners extract the rules of the grammar from learned instances and the rules become grammaticalized for productive use by the parser, thus eliciting a P600 effect. Thus, when it comes to understanding the acquisition of grammatical processing, the recruitment of lexico-semantic processing streams has been thought of as a temporary steppingstone to the native-like use of a distinct grammatical processing stream, reflected in the P600 response to grammatical errors. Moreover, the rate of this transition is not homogenous across learners. McLaughlin and colleagues found that a subset of learners had already begun to show a P600 effect halfway through instruction (16 weeks) whereas the remainder continued to response with an N400 effect to grammatical errors, suggesting they were slower at learning the L2 rule. That some learners made this transition faster than others provided empirical support for a commonly made observation that some individuals lean an L2 at a faster rate than other leaners.

1.3 Individual differences in grammar processing

Data from both L1 and L2 comprehension literature seem to suggest that neural responses to grammatical violations that do not approach the canonical P600 effect reflect subpar syntactic processing. However, this belief that the P600 response is an index of effective grammar processing rests on the assumption that it reflects the neurocognitive "end state" for all typical, proficient comprehenders. This assumption has been challenged in recent years as traditional group analyses have given way to studies of individual differences in grammar comprehension.

Individual differences research has instead revealed marked variability in the way that native speakers process the same grammatical errors (Kim et al., 2018; Nakano et al., 2010; O'Rourke & Colflesh, 2015; Osterhout, 1997; Pakulak & Neville, 2010; Qi et al., 2017; Tanner, 2019; Tanner & Van Hell, 2014). First, the magnitude of the P600 effect is not homogeneous among native speakers (Pakulak & Neville, 2010; Tanner, 2019; Tanner & Van Hell, 2014). Although there is underlying variability to any average, the fact that the P600 effect is attenuated in some healthy, adult comprehenders compared to others suggests that the same grammatical errors do not violate syntactic expectations in similar ways across native speakers. Second, syntactic anomalies do not always elicit the canonical P600 effect in all comprehenders. An early study of garden path sentences demonstrated that while most L1 English speakers elicited a P600 response to the temporarily anomalous word (e.g., *sank*) in sentences like "*The boat sailed down the river* <u>*sank*</u>", a non-trivial proportion showed an N400 effect instead (Osterhout, 1997). For these participants, it appeared that *sank* violated semantic expectations rather than syntactic ones: the interpretation of "*The boat sailed down the river*" as a completed action may have led the completion *sank* to be an unexpected additional action. Tanner and van Hell (2014) further demonstrated that this is not a categorical distinction, but rather that individuals fall along a continuum with regard to whether a grammatical error violates form- or meaning-based predictions. The authors designed the Response Dominance Index (RDI) to measure the degree to which N400 and P600 effects dominated a participant's neural response, where more negative values reflect N400-dominance and more positive value reflect P600-dominance. RDI scores in response to subject-verb agreement errors (e.g., *The clerk at the clothing store* <u>*was/*were*</u> *severely underpaid*) revealed that individuals spanned the entirety of this continuum, from being

N400-dominant, to biphasic (N400 and P600 of similar magnitude), to P600-dominant. Moreover, responses to subject-verb agreement errors were positively correlated with responses to verb tense agreement errors (e.g., *The crime rate was increasing/*increase despite the growing police force*): individuals who were N400-dominant or P600-dominant in one error showed similar ERP dominance patterns in the other.

Several variables have been proposed to explain why grammatical errors elicit attenuated P600 responses or N400 responses in some speakers. Capacity theories of comprehension have previously attributed variability in garden path processing to differences in working memory (WM): individuals who are capable of maintaining multiple parses in working memory recover more quickly from the ambiguity compared to individuals who commit to one structure and must re-parse the sentence (Just & Carpenter, 1992; King & Just, 1991). Individuals with lower WM capacity might thus be expected to respond with larger P600s. Individuals with lower WM capacity did show larger P600 effects at the disambiguating word that resolved a syntactic ambiguity compared to high span participants (Friederici et al., 1998), but the larger P600 effect was interpreted as increased processing effort at the disambiguating word due to low span comprehenders inability to commit to any parse until disambiguating information came in. However, the opposite interpretation has been made to account for the effect that verbal WM has on variability in semantic error processing involving animacy violations, such as "*The box was ?biting*[2] *the man*" (Kim et al., 2018; Nakano et al., 2010). Individuals with better verbal WM may be able to access the animacy feature of *box* more quickly and plan a passive construction early on because inanimate nouns occur less frequently in subject positions. Upon encountering

---

[2] A question mark (?) is linguistic notation for semantic anomaly. When a questions mark appears next to a word, it signifies that the word is considered semantically anomalous given the context.

an active construction after all, high verbal WM individuals need to reprocess their planned syntactic structure and thus elicit larger P600 magnitudes. Regardless of which interpretation is best suited, there is mounting evidence that differences in ERP responses to grammatical violations are not well accounted for by WM limitations (O'Rourke & Colflesh, 2015; Pakulak & Neville, 2010; Tanner, 2019; Tanner & Van Hell, 2014). For instance, even simple phrase-structure violations written for comprehension by 3-year-olds (e.g., *Timmy can ride the horse at my *his farm*) elicited variable ERP responses in English-speaking adults (Pakulak & Neville, 2010). Pakulak and Neville (2010) recruited participants from the general population in order to capture variability beyond the common university setting and found that English proficiency, not verbal working memory, predicted speakers' neural responses. Compared to high proficiency individuals, low proficiency participants showed less pronounced P600 effects in addition to N400-like negativities in anterior electrode locations (for evidence that anterior negativities are artifacts of averaging N400 and P600 effects, see Tanner, 2015). These data align well with related findings in the L2 grammar processing literature where ERP variability and proficiency are tightly linked.

Although the relationship between proficiency and language processing may appear to be the most parsimonious explanation for individual variability in comprehension, processing differences have still been observed even among highly proficient native speakers (Tanner, 2019; Tanner & Van Hell, 2014). Tanner (2019) examined ERP responses to simple grammatical violations in a large cohort of native English speakers who completed several measures of language experience and working memory skill. Latent variables for each construct were extract from these tasks and used to predict N400 and P600 magnitudes across two types of subject-verb agreement errors and reading tasks. RDI values were highly positively correlated across violations in both reading tasks, but neither language experience nor working memory predicted grammar processing variability. Thus, semantic and syntactic processing streams dominate an individual's grammar processing in a way that is consistent and that cannot be accounted for by differences in language experience or working memory. This led the author to conclude that the RDI reflects a stable trait of an individuals' grammatical agreement processing rather than a product of related cognitive and experience domains.

Data from individual differences research thus indicate that far from being the "end state" for all native speakers, the processing streams that elicit the P600 effect can in fact compete with the processing streams that elicit the N400 effect for dominance even among highly educated and proficient native language comprehenders (Tanner, 2019; Tanner & Van Hell, 2014). As previously illustrated, grammatical errors violate syntactic expectations to a greater degree in some people and semantic expectations in others. Averaging data from these individuals masks the natural variability evident among native speakers (Tanner, 2015) and creates an artificial, one-size-fits-all target for proficient L1 comprehenders and L2 learners alike. Given these observations, it should no longer be a foregone conclusion that an N400-dominant response to a grammatical violation necessarily reflects atypical processing, nor that L2 learners who elicit N400 responses to syntactic violations are necessarily delayed in their acquisition. For instance, variability in native Spanish learners' processing of grammatical errors in English was not explained by differences in English proficiency (Tanner et al., 2014). Relatedly, it is likely that some proportion of McLaughlin et al.'s (2010) N400 group were no less proficient in the French rule than their P600 counterparts when they were tested halfway through instruction. If the N400 response was suboptimal, an inverse relationship between N400 magnitude and proficiency would be expected. Yet no such correlation was found. As for the acquisition of L2 rules that are not instantiated in the L1, the absence of a P600 effect need not necessarily mean that L1-L2 dissimilarity precludes successful acquisition or native-like use of unfamiliar grammatical rules. In fact, the use of the P600 a metric of linguistic success suffers from the same issue as many other measures that are used to test whether new grammatical rules can be acquired after the critical period of L1 acquisition. These measures rely on native speaker norms that are often gathered from nonrepresentative samples and do not to capture the full range of variability in the baseline against which L2 learners are compared. As a result, they are fail to detect nativelike performance in many learners (Andringa, 2014).

1.4 Alternative explanations of ERP effects and their variability

The individual differences research discussed thus far suggests that variability in syntactic processing is a stable feature of comprehenders that is explained only in part by differences in working memory and language experience. Whether ERP patterns of grammar comprehension indeed reflect an individual trait warrants further investigation, especially with regard to the factors that can account for the source of this feature. However, any alternative explanations need to account for two findings. The first is that L1 and L2 systems share similar cognitive (Kroll & Sunderman, 2003; Sunderman & Kroll, 2006) and neurological mechanisms (Abutalebi, 2008; Kotz, 2009; Perani & Abutalebi, 2005; Sulpizio et al., 2020). If variability in grammatical processing is a trait of the comprehender then it should manifest in L1 and L2 processing in similar ways through these shared mechanisms. Thus, an alternative explanation should be able to account for differences in both L1 language processing and L2 learning. Whereas language experience has been shown to influence individual differences in L1 and L2 processing, previous research has not considered the degree to which working memory ability explains variability in ERP responses among L2 learners. Moreover, factors that have been central for explaining variability in L2 grammar acquisition, such as language aptitude, learning strategy, and motivation (Dörnyei & Skehan, 2003; Naiman, 1996; Robinson, 2002; Tanner et al., 2014), have not yet been related to ERP individual differences in L1 grammar processing.

Second, alternative explanations should be able to account for the finding that the P600 effect can also be elicited by various syntactically well-formed sentences that contain errors of meaning (Chow & Phillips, 2013; DeLong et al., 2014; Hoeks et al., 2004; Kim & Sikos, 2011; Kolk et al., 2003; Kuperberg et al., 2006; Osterhout & Mobley, 1995; van de Meerendonk et al., 2010; van Herten et al., 2005; Van Petten & Luka, 2012). For instance, although sentences containing animacy violations (e.g., "*The box was ?<u>biting</u> the man*") or thematic role reversals (e.g., "*The fox that hunted the ?<u>poachers</u>…*") are grammatical, their errors result in P600s. Given that the same ERP component can be elicited by both syntactic and semantic violations, any arguments made with regard to the nature or the source of variability in syntactic P600 magnitude should also extend to the semantic P600. If individual differences in grammar processing are believed to be a trait of the comprehender, this trait should also manifest in the context of semantic error processing. Thus, any variable is proposed to explain P600 variability in syntactic error processing should also be able to explain why the P600 occurs in the case of grammatical stimuli.

Of the various grammatical stimuli that result in a P600 effect, there is a manipulation of particular interest that may provide one such explanation for individual differences in grammar processing that meets the above stated requirements. The semantic attraction violation occurs in syntactically well-formed sentences such as "*The hearty meal was <u>devouring</u>…*" and elicits P600 responses in native speakers (Kim et al., 2018; Kim & Osterhout, 2005). This type of sentence presents an interesting case because the nature of its anomaly is ambiguous: it can be perceived as violating either syntactic or semantic expectations depending on the way linguistic cues inform interpretation. For example, *meal* is in the subject position of the sentence, which is a strong syntactic cue that it should be assigned the thematic role of agent (i.e., who or what is doing the action) of the main verb following the auxiliary *was*. If *meal* is the agent, then the verb is expected to end in -*ing* following the rules of the progressive construction signaled by *was*. However, due to animacy constraints, inanimate objects such as *meal* cannot perform active actions such as *devour*. The verb *devouring* would thus match the expectations of the syntactic cues (i.e., end in -*ing*) but be interpreted as semantically inappropriate for the inanimate agent *meal*, resulting in an N400 effect. That is, perhaps "*The hearty meal was cooking…*" would be more appropriate. Nevertheless, the authors did not observe an N400 as they predicted. Results instead suggested that participants were highly influenced by the semantic relationship between the subject *meal* and the verb *devour*. This semantic attraction created a highly plausible discourse model related to the devouring of meals. Its influence was so strong that it led to *meal* being assigned the thematic role of theme (i.e., to whom or to what the action is being done) for the verb *devour*, overriding the available syntactic cues for the agent role. The verb *devouring* was thus semantically plausible but ungrammatical because verbs in passive constructions (where the theme appears in the subject position) require -*ed* endings (e.g., "*The hearty meal was devoured…*"), resulting in a P600 effect.

The reported group data from Kim and Osterhout (2005) suggest that strong semantic cues can cause native speakers to interpret syntactically well-formed sentences as being syntactically anomalous. These data are contrary to predictions of syntax-first models of language processing, which propose that syntactic processing takes place before semantic processing and therefore drives semantic interpretation (Ferreira & Clifton, 1986; Frazier &

Clifton, 1997; Frazier & Rayner, 1982; Friederici, 2002). Instead, semantic and syntactic information interact during sentence comprehension as input is combined for interpretation. Depending on the construction and context, certain linguistic information can guide this process to a greater degree. For example, when the semantic attraction between the subject noun and the verb was removed (e.g., "*The dusty tabletops were* <u>*devouring*</u>…"), participants showed the predicted N400 effects (Kim et al., 2018; Kim & Osterhout, 2005). The weaker semantic relationship between *tabletops* and *devour* was not enough to drive comprehension in this condition.

If the relative contribution of semantic or syntactic information can influence processing in a group of comprehenders, then it is plausible that these linguistic cues may also compete to different degrees at the level of the individual. Perhaps, then, variability in RDI values and ERP response patterns observed in the individual differences research reflects the relative influence of semantic or syntactic cues on sentence comprehension within an individual. I analyzed data from Kim and Osterhout (2005) by participant and found the same individual differences previously observed in other linguistic constructions, where a sizeable portion of individuals showed N400 effects in contrast to the P600 responses demonstrated in the group average. For these individuals, the semantic attraction between *meal* and *devour* did not override the syntactic cues and participants interpreted *devouring* as a semantic anomaly, as the authors predicted the group would do.

Kim and colleagues (2018) also detected individual differences in semantic attraction sentence processing, which they were able to attribute to variability in verbal WM capacity. Individuals with high verbal WM capacity demonstrated P600 effects, whereas low capacity individuals showed N400 effects. No relationships with non-verbal WM capacity or linguistic knowledge were found. The authors suggested that high verbal WM capacity individuals are more likely to attempt structural reanalysis because they are capable of maintaining the syntactic representation in mind as they reinterpret the sentence. Low capacity individuals, on the other hand, opt instead for looking for a meaning that is more appropriate to the context. However, it is not clear that the search for a context-appropriate interpretation would not also depend to some degree on working memory. Neither is it evident why high WM capacity would necessitate engagement of syntactic processes over semantic processes. In fact, a related study did not find this correlation between variability in P600 responses to semantic anomalies and verbal WM (Kos et al., 2012). Given that there is also limited evidence for its role in L1 and L2 ERP grammar processing variability, WM capacity does not appear thus far to be an adequate candidate for explaining the trait of individual language processing differences.

This dissertation therefore proposes that individual differences in grammatical processing may result from variability in the degree to which syntactic information influences ongoing interpretation within a comprehender. That is, syntactic and semantic cues compete within an individual to different degrees. Under this proposal, individuals who show larger P600 magnitudes or more positive RDI values give more weight to information provided by grammar cues than individuals who show larger N400 magnitudes or more negative RDI values. Further, L2 learners who transition to a P600 response may also be influenced by syntactic information to a greater degree than L2 learners who are slower to transition or who show N400 responses to grammar violations instead. Which cues ultimately drive interpretation within an individual will depend on one of two hypothesized explanatory variables, which will be explored in greater detail in the following sections.

The first hypothesized explanation the current experiments will test will be termed the

Top-Down Prediction account, which proposes that individuals vary with regard to the type of linguistic information they attend to in order to build prediction of upcoming input. Under this account, individuals who rely to a greater degree on grammatical cues build stronger top-down syntactic predictions. The stronger these syntactically based predictions are, the more sensitive individuals are to grammatical violations and the larger their P600 magnitudes get. Comprehenders who show N400 effects to grammatical violations, on the other hand, may instead be adjusting their semantic predictions. For example, in a subject-verb agreement error such as "*The roses *grows in the garden*" they might update their semantic predictions to reflect a context where what was growing in the garden was singular rather than plural. In the context of the semantic attraction data, syntax-focused individuals would be those participants who were not affected by the semantic relationship between *meal* and *devour* and showed N400 responses to *devouring*, which matched their syntactic predictions but were interpreted as semantically anomalous. In contrast, the pull of the semantic attraction would be especially strong in semantic-focused individuals. These comprehenders would thus respond with a P600 effect to *devouring*, which matches the discourse model set up by the relationship between *meal* and *devour* but is ungrammatical.

The second hypothesized explanation will be termed the Syntactic Sensitivity account, which proposes that individuals vary with regard to their sensitivity to grammatical form, or grammar aptitude. Under this account, individuals who have greater grammar aptitude have better awareness and knowledge of syntactic structure and information, and thus have more responsive grammar analytic processes. Upon encountering a grammatical error, grammar processing streams are engaged to a greater degree and this manifests as a larger P600 response. Comprehenders who have poorer grammar aptitude, on the other hand, are generally less sensitive to syntactic errors and rely less on syntactic processing streams to resolve syntactic violations. These individuals may show smaller P600 effects to "*The roses *grows in the garden*" because the meaning is still generally intact and not much syntactic reanalysis is necessary in order to proceed with comprehension. In contrast with the Top-Down Prediction account, individuals who show P600 effects to semantic attraction violations would instead be those individuals who are more sensitive to grammatical form. Similar to Kim et al.'s (2018) interpretation, these comprehenders would prefer to resolve the ambiguity using syntactic processing streams. On the other hand, individuals who are less sensitive to grammar would instead interpret *devouring* as a semantic violation due to the animacy restrictions of *meal* and demonstrate an N400 response.

Evidence for each of these hypotheses will now be examined in turn with regard to both L1 and L2 grammar processing literature. Predictions for the present study under each of these hypotheses will then be discussed.

1.4.1 *The Top-Down Prediction account*

The Top-Down Prediction account proposes that individuals vary with regard to the type of linguistic information they attend to that informs their predictions of upcoming input. This hypothesis emerges from the fact that traditional interpretations of the N400 as a metric of semantic analysis and integration and the P600 as a metric syntactic reanalysis and repair have had to contend with the above described findings that N400 effects are sometimes found in response to syntactic errors and P600 effects are something found in response to semantic errors. There is instead an emerging trend to interpret these ERP components as indexing predictive processes necessary for effective comprehension. This view already aligns well with the established literature on changes in N400 magnitude related to contextual expectancy. Recall that the N400 is negatively correlated with cloze probability and is larger for semantically unexpected but plausible and semantically implausible continuations (e.g., "*The bill was due at the end of the* *__hour__*/?*__paint__*") relative to expected continuations (e.g., "*The bill was due at the end of the* *__month__*") (Federmeier & Kutas, 1999; Kutas & Hillyard, 1984). This behavior of the N400 provides evidence that the comprehension system uses contextual information to actively predict semantic features of incoming input, even pre-activating individual words in a graded manner based on the likelihood of their continuation. For example, N400 magnitudes were not only larger to the less expected continuation *airplane* in "*The day was breezy so the boy went outside to fly* *__a kite__*/*__an airplane__*," but a difference in magnitude was already apparent at the article *an* (DeLong et al., 2005; Delong et al., 2011). The context led to a strong prediction for the ending *kite* and thus its matching article *a*, which facilitated processing compared to the unexpected article *an*. Thus, the N400 can be understood as indexing the degree to which semantic predictions are fulfilled, such that greater N400 magnitudes reflect a greater mismatch between predictions and the input.

The P600 has also been found in response to sentence continuations that violate strong contextual constraints, such as in the case of animacy violations, as already described (e.g., "*The box was* ***__*biting*__ *the man*"), but also in the case of implausible sentence continuations (e.g., "*The bill was due at the end of the* ?*__paint__*") (Federmeier et al., 2007; Kuperberg et al., 2003, 2006, 2007, 2020; Pijnacker et al., 2010; van de Meerendonk et al., 2010; Van Petten & Luka, 2012). In this latter example, the implausible continuation does not violate animacy restrictions or thematic role assignment: it is simply an impossible interpretation given the context. This is notable because it demonstrates that the P600 does not only come online for violations where semantic and syntactic information interacts (e.g., word position influences thematic role assignment, and vise versa). Instead, it also appears to be sensitive to conflict between top-down predictions and the bottom-up input like the N400. However, the P600 may instead be responsible for monitoring predictions related to structure and form that underlie relationships between events. An implausible continuation would then fail to match expectations regarding the underlying structure of the event predicted by the grammatical form of the sentence.

The idea that the P600 may track predictions related to structural relations is compatible with the established grammar processing literature as well. For example, P600 magnitudes were larger for syntactic violations following modal verbs like *should* (e.g., "*The sheep should *<u>grazing</u> in the pasture*") compared to violations following auxiliary verbs like *were* (e.g., "*The sheep were *<u>graze</u> in the pasture*") (Mehravari et al., 2015). As per the traditional interpretation of the P600 as a syntax-specific response, the authors attributed this difference to the morphological complexity of *grazing*, having two morphemes (i.e., *graze* + *-ing*), relative to the simple morphology of *graze*. However, the alternative view of the P600 as an index of conflict with structural predictions would argue instead that *grazing* mismatched predictions to a greater degree than *graze*. This is plausible because modal verbs like *should* are always followed by an infinitival verb form (e.g., *should graze*/*sleep*/*run*, or *should be/have…*, etc.) whereas auxiliary verbs like *were* need not be followed by a verb at all (i.e., they can also be proceeded by adjectives, adverbs, prepositions, etc.). Thus, the ungrammatical verb in the modal verb condition would have violated structural predictions to a much greater degree than in the auxiliary verb condition.

This interpretation of the P600 is also well suited to explain its occurrence in violations music (Patel, 2003; Patel et al., 1998) and arithmetic (Fisher et al., 2010; Núñez-Peña & Honrubia-Serrano, 2004), two domains that are also highly dependent on structural relations between elements for coherent meaning. Not only does the presence of the P600 in these domains provide evidence against a syntax-specific role for this component, but these data provide further support that the P600 is sensitive to violations of strong contextual constraints and is attenuated in a graded manner by the degree of the violation. Patel and colleagues (1998) showed that P600 magnitude increased as the final chord in a sequence increased in musical distance from the key of the sequence (e.g., in-key vs. near-key vs. distant-key). P600 magnitude was also found to increase as the final number in an arithmetic series increased in distance from the correct series continuation (e.g., 1, 4, 7, 10 followed by 13 vs. 15 vs. 20) (Núñez-Peña & Honrubia-Serrano, 2004). In both cases, stimuli that were furthest from the predicted structural relations elicited P600 effects that reflected a great degree of mismatch from the expected structural continuations.

1.4.1.1 Prediction in the communication model of comprehension

Kuperberg and colleagues (2020) have further qualified the role of the N400 and the P600 in light of a proposed hierarchical language processing model in which higher levels of representation inform top-down predictions for information processed at lower levels of the hierarchy. When predictions and input mismatch, information is sent up from the lower levels to provide feedback to and update the higher-level models that inform predictions (Kuperberg, 2007; Kuperberg & Jaeger, 2016). They propose that the comprehender builds a communication model that is informed by both linguistic and non-linguistic information, such as available knowledge about the context and the communicator. The communication model is composed of three levels. At the highest level is the situation model, which maintains a representation of the ongoing meaning being built from information about the events and actions taking place. This situation model drives predictions about the event model, the next level of representation. The event level maintains information about the types of events that are plausible given the situation model as well as the events currently under discussion. At the lowest level is the semantic feature model, which holds more fine-grained representations of the semantic properties of incoming words and features of words that are compatible with the events being maintained at the level of the event model.

In the context of this communication model, they propose that the N400 indexes the degree to which incoming words match the features at the semantic feature level. The closer the match, the smaller the N400 magnitude. The P600, on the other hand, tracks representations at the event level by checking the input against the situation model constraints. The smaller the mismatch, the smaller the P600 magnitude. Finally, an additional ERP component, the late frontal positivity, is responsible for updating the situation model when representations at the event level don't quite match the expectations of the high-level situation model. Like the P600, the frontal positivity is a large, positive-going waveform that occurs between 500 and 1000 ms, but it is maximally distributed across frontal and pre-frontal electrode sites rather than posterior ones. The frontal positivity occurs when a context constrains strongly for a highly expected continuation but is followed by an unexpected but plausible one (e.g., "*The bill as due at the end of the <u>hour</u>*") (Delong et al., 2011; Kuperberg et al., 2007, 2020; Sitnikova et al., 2008; van de Meerendonk et al., 2009, 2010; Van Petten & Luka, 2012). These continuations may not fully match the events predicted by the situation model but, being plausible, can be integrated into the situation model, which will be updated to reflect the new meaning.

1.4.1.2 Prediction in neural oscillations

Whether the N400 and P600 components are responsible for the specific predictive processes outlined in Kuperberg et al.'s (2020) communication model or not, the data described thus far support the conceptualization of these components as indices of linguistic prediction at different levels of linguistic representation. Data from neural oscillations provide further converging evidence for this view. Recall that ERPs are produced by averaging the EEG signal to reveal fluctuations in voltage over time. However, an electrical signal can alternatively be decomposed into its component frequencies to reveal how prevalent certain frequencies are within the electrical signal at a given time. When the neural EEG signal is decomposed in this way, its component frequencies are believed to reflect oscillatory activity of populations of neurons in the cortex that synchronize and desynchronize as cognitive processes are performed. Groups of neurons within a single functional network will couple by firing synchronously at a particular frequency. This synchronous oscillatory firing both entrains functional networks of neuronal populations involved in a given function, allowing for different information to be represented at different frequencies, and also helps bind different levels of information being processed as part of one function (Gray et al., 1989; König & Schillen, 1991). When information no longer needs to be represented or shared across a network, the network desynchronizes. These neural oscillations can thus be thought of as the foundation of the brain’s language.

Enumerable cognitive processes have been related to synchronization and desynchronization of oscillatory activity at specific frequency bands, but two frequencies are particularly important for understanding prediction in the brain. When sensory input from the thalamus enters layer 4 of the cortex, it travels to superficial layers (L2/L3) that connect to other cortical areas via feedforward connections for bottom-up processing, and information is then sent back from L2/L3 to the deep layers (L5/L6) for top-down influence via cortico-thalamic feedback projections (Bastos et al., 2012; Douglas et al., 1989; Douglas & Martin, 1991; Gilbert & Wiesel, 1983). There is evidence that feedforward signals sent from superficial layers of the cortical hierarchy are typically transmitted by higher frequency oscillations in the gamma range (30-100 Hz), while feedback signals sent from to the deeper layers of the cortical hierarchy are transmitted by lower frequency oscillations such as those in the beta range (12-30 Hz) (Bastos et al., 2012; Wang, 2010). This organization of the cortical hierarchy and its pattens of neuronal processing provide a physiological basis for predictive coding in the brain, where different neuronal populations are involved in different computations related to prediction (Bastos et al., 2012; Friston, 2010; Friston et al., 2015). Bastos et al. (2012) propose that beta oscillations code top-down information that provides contextual predictions at lower levels of processing. Incoming bottom-up information is coded by gamma oscillations. This high frequency oscillation allows for fast propagation of information for higher-level processing. Over time, bottom-up information starts to build context, which is fed back to inform processing of bottom-up input via slower beta oscillations. When new bottom-up input arrives, the gamma frequency also communicates prediction errors, or the degree of mismatch between the prediction and the input. This high frequency allows for fast propagation of error signals, allowing for rapid adjustment of the contextual framework at upper levels of processing and updating of predictions encoded by beta oscillations.

Studies of neural oscillations during sentence processing provide evidence for the predictive coding hypothesis in the domain of language comprehension. Beta power (i.e., amount of oscillatory activity in the 12-30 Hz range) has been shown to increase over the course of well-formed sentences (Bastiaansen et al., 2010; Bastiaansen & Hagoort, 2015; Pefkou et al., 2017; Segaert et al., 2018) and decreases at the occurrence of syntactic or semantic violations (Bastiaansen et al., 2010; Davidson & Indefrey, 2007; Kielar et al., 2014, 2019; Lewis et al., 2016; Li et al., 2014; Luo et al., 2010; Pérez et al., 2012; Vignali et al., 2016; Wang, Jensen, et al., 2012) and at the end of speech turns (Magyari et al., 2014). Increases in beta power have also been found in contexts where the top-down representation needs to be maintained, such as in syntactically complex contexts (e.g. “*The mouse that the cat chased ran away*”) relative to syntactically simpler ones (e.g., “*The cat that chased the mouse ran away*” (Bastiaansen & Hagoort, 2006; Meyer et al., 2013; Weiss et al., 2005), or when the end of speech turn is not predictable (Magyari et al., 2014). These data suggest that the beta oscillation is responsible for building, updating, and maintaining a neurocognitive network responsible for the ongoing sentence-level meaning under construction, or the cognitive set (Bressler & Richter, 2015; Engel & Fries, 2010; Lewis & Bastiaansen, 2015). When input fails to meet the expectations of the cognitive set or when the cognitive set is no longer needed, such as at the end of a speech turn, beta decreases.

Data from studies of speech perception provide evidence that gamma frequencies are involved in propagation of prediction errors (i.e., the degree of mismatch between predictions and input) during language processing as well (Arnal et al., 2011; Arnal & Giraud, 2012; Fontolan et al., 2014; Giraud & Poeppel, 2012). Lewis and Bastiaansen (2015) also propose that gamma oscillations in the lower part of the gamma frequency range may also be responsible for suppressing competing lexical representations, such as when the comprehender encounters a semantic violation instead of a highly expected word (Hald et al., 2006; Penolazzi et al., 2009; Rommers et al., 2012; Wang, Zhu, et al., 2012; Weiss & Mueller, 2003). They further argue that gamma is broadly responsible for representing and communicating the degree of mismatch between top-down predictions and bottom-up linguistic input (Herrmann et al., 2004).

Thus, beta and gamma frequency ranges are believed to index different facets of linguistic prediction. However, whereas beta changes have been found to occur in both syntactic and semantic violations, gamma changes have primarily been found in semantic violations and speech perception errors and so may be limited to local error propagation at lower hierarchical levels of linguistic feature processing. Moreover, rarely have beta and gamma changes been found to occur in the same stimuli, as one might expect if top-down and bottom-up processes inform each other within a hierarchical framework (for a review, see Prystauka & Lewis, 2019). Speculation as to whether beta and gamma reflect separate but complementary processes or whether one process wins out over the other depending on the type of stimulus is beyond the scope of this dissertation. Nevertheless, it is notable that the same type of stimuli that elicit N400 and P600 effects also show decreases in beta at the point of the violation, providing evidence that these ERP components may be involved in building top-down predictions. In fact, recent work showed that the magnitude of the P600 effect and the degree of beta desynchronization were positively correlated across stimuli containing grammatical violations (Schneider & Maguire, 2018). The more a stimulus violated syntactic expectations, the more the maintenance of the cognitive net that supported those expectations was interrupted. Not only does is this data support the Top-Down Prediction account, it provides a secondary measure with which to test its validity. If individual differences in language processing are due to differences in the type of information comprehenders use to build predictions, then syntactic violations would not only elicit larger P600 effects in individuals who attend more to syntactic cues, but they would also cause greater decreases in beta power.

1.4.1.3 Prediction in L2 learning

The deployment, maintenance, and adjustment of top-down linguistic predictions is not only relevant for effective language comprehension but for language learning, too. In early stages of learning, individuals accumulate information about statistical regularities in their linguistic input over a series of exposures. As learning progresses, these statistical regularities allow learners to develop predictions about what they are likely to encounter given. For instance, an English learner will encounter the co-occurrence of the plural *-s* ending on nouns that follow *these* but not *this*. Eventually, they will begin to predict the presence of an *-s* ending when they process the plural demonstrative *these*. However, whether "eventually" occurs within a short period of time or a long period of time could depend greatly on how quickly a learner begins making predictions. Making predictions, even wrong ones, is essential for learning because predictions are always followed by prediction error, which allows those predictions to adapt. Moreover, individuals who had greater beta synchronization were also better at learning grammar rules that specified fixed word order but not grammar rules that have flexible word order (Cross et al., 2020). The fixed word order rule allowed learners to build predictions faster within the short training period. Individuals who built stronger predictions, as indexed by greater beta synchronization, were better at acquiring the implicitly taught rule and the degree of their beta synchronization predicted subsequent performance.

In fact, learning might even take place more quickly in the case of incorrect predictions because incorrect predictions will result in large prediction errors and large prediction errors lead to better learning. Thus, individuals who do not wait until they have sufficient statistical information to make correct predictions may be faster to acquire linguistic rules and able to process language more efficiently. For example, individuals who performed better at prediction-based learning were also better at comprehending complex sentences (Misyak et al., 2010). Evidence suggests that sensitivity to errors can facilitate learning that happens quite quickly, even within the span of a couple of hours (Bastarrika & Davidson, 2017; Batterink & Neville, 2014; Bultena et al., 2017; Davidson & Indefrey, 2009, 2011). Davidson and Indefrey (2009) taught novice L2 German learners a grammatical gender agreement rule over the course of two experimental sessions. During the training phase, ERPs were recorded while participants received feedback about incorrect grammaticality judgments. The magnitude of the error-related negativity, an ERP response that occurs after such feedback, predicted the degree to which learners' accuracy scores improved from pre- to post-training, suggesting that the degree of a learner's prediction error modulated subsequent linguistic performance. Data from these training studies underscore the role of feedback for stabilizing L2 representations because it ensures error detection and thus learning.

Even proficient native speakers can be taught to adjust predictions within a single experimental session. Native German speakers who were exposed to German nouns that were assigned incorrect genders subsequently stopped using gender information to make predictions during comprehension (Hopp, 2016). Having noticed that gender information in this context was not aligned with their knowledge of the language, German speakers adapted their own predictions. Thus, attention to linguistic information is a necessary step in the development and adaptation of linguistic expectations. Batterink and Neville (2014) showed better syntactic learning in L2 French learners who spontaneously attended information about word class during sentence processing, allowing them to acquire a simple morphosyntactic rule in French within the span of two hours. These data are directly relevant to the application of the Top-Down Prediction account to the context to L2 acquisition. Under this account, individuals who attend more to syntactic information and build stronger syntactic predictions will also be predicted to acquire L2 grammatical rules more quickly than individuals who attend more to semantic information, assuming differences in proficiency and experience with the language are accounted for. Moreover, given that linguistic predictions depend to a great degree on attention and need not be restricted by existing knowledge of the native language, this prediction would also hold in the absence of L1-L2 similarity. That is, even acquisition of L2 rules that are considered complex or difficult to acquire because they are absent in the L1 should be made possible and accomplished with greater ease by individuals who attend more to syntactic information.

1.4.1.4 Implications for individual differences

To reiterate, the Top-Down Prediction account of individual differences in grammar processing proposes that individuals vary with regard to the type of linguistic information they attend to in order to build predictions of upcoming input. The evidence presented thus far shows that prediction is an integral part of language processing and language learning that has been substantiated in both ERP and neural oscillation research, albeit primarily at the group level. If the N400 and P600 are indexes of linguistic predictive properties, then whether a grammatical violation results in an N400 response in one individual and a P600 response in another may instead signal differences in the way that individuals predict rather than a deficiency in the grammar processing system. A study of deaf and hearing individuals provides some preliminary evidence for this inference. ERP responses to double violations, which contained simultaneously semantically- and syntactically-violated errors (e.g., "*The huge houses still */?<u>listens</u> to my aunt*") were recorded in deaf and hearing individuals who were highly skilled readers (Mehravari et al., 2017). Reading skill in hearing individuals was predicted by magnitude of the P600 effect in responses to the double violation but reading skill in deaf individuals was predicted by magnitude of the N400 effect in response to the double violation. Not only did the two groups pay attention to different kinds of linguistic information when processing the same error, both groups of individuals were capable of achieving strong language skills regardless of the type of linguistic information they attend to. These data suggest that variability in ERP responses could provide a window into predictive processes at the level of the individual as well.

1.4.1.4.1 Predictions for individual differences in L1 processing

If N400 effects are assumed to index semantic predictions and P600 magnitudes are assumed to index syntactic predictions, then it is proposed that individuals who show larger N400 effects rely more strongly on semantic information for building predictions and individuals who show larger P600 effects rely more strongly on syntactic information for building predictions. If this is an appropriate characterization of an individual's grammatical comprehension processes, then the predilection to attend to different linguistic information should be observed across different linguistic stimuli within an individual. Individual differences in ERP response dominance have already been observed across syntactic violations (Tanner, 2019; Tanner & Van Hell, 2014), but these violations were fairly similar, as they all contained errors of agreement: three of the four conditions contained subject-verb agreement violations and one contained a verb tense agreement violation. Form and meaning interact in rules of agreement, which make them a good candidate for testing the degree to which semantic or syntactic predictions are violated (e.g., proper agreement in "*The rose* *grows*…" relies on semantic knowledge that the *rose* is singular and not plural). However, if response dominance patterns reflect a grammar processing trait of the individual, then similar dominance patterns should also be observed in syntactic violations where form and meaning interact to greater or lesser degrees. For example, P600-dominant individuals might show even more positive RDI values (i.e., more P600-dominant responses) in response to strictly of form-based errors (e.g., word-order/phrase structure violations: "*Timmy can ride the horse on my *his farm*"). On the other hand, N400-dominant individuals might show even more negative RDI values (i.e., more

N400-dominant responses) in response to more semantically-involved errors (e.g., subject-verb agreement errors where the subject is semantically plural but syntactically singular, such as *committee*: “*The committee *were meeting later that evening*”).

Given that prediction permeates all aspects of language comprehension, not just those related to grammar processing, it is further posited that ERP dominance patterns elicited by syntactic errors should also be related to response dominance patterns elicited by semantic errors. That is, syntactic and semantic errors should both result in more P600-dominant responses in individuals who build strong syntactic predictions and more N400-dominant responses in individuals who build strong semantic predictions. Furthermore, this differential sensitivity to semantic violations should apply both for implausible semantic continuations as well as for unexpected but plausible semantic continuations. If an individual attends more to syntactic information and builds stronger syntactic predictions, then that individual’s predictions will be less focused on the fit of a semantic continuation given the context and more on whether the continuation meets structural expectations. In other words, syntax-focused individuals would care more that a determiner like *the* is appropriately followed by a noun than the degree to which that noun fits the contextual constraints of the sentence. Thus, individuals who are determined to be P600-dominant in response to syntactic errors would be expected to show smaller N400 effects to unexpected but plausible or implausible semantic continuations than individuals determined to be N400-dominant.

Finally, if N400 and P600 ERP components indeed index prediction-related processes, then ERP responses to syntactic or semantic violations should be related to degree of beta desynchronization as a result of those violations. Recall that beta desynchronization occurs when the cognitive network related to top-down prediction is no longer appropriate or needs to be adjust given the input. The more beta desynchronizes, the more the maintenance of the ongoing cognitive set is reduced. If one individual shows greater beta desynchronization in response to a violation than another, then it can be said that the error violated that individual's prediction to a greater degree. Thus, RDI values and beta power manifest two separate but complementary features of prediction: RDI values reveal the type of linguistic prediction being made, whereas beta desynchronization reveals the strength of that prediction. In this way, P600-dominant individuals who build stronger syntactic predictions would be expected to have larger decreases in beta in response to grammatical violations compared to N400-dominant individuals. These P600-dominant individuals would also be expected to have smaller decreases in beta in response to semantic violations compared to N400-dominant individuals.

1.4.1.4.2 Predictions for individual differences in L2 learning

Given that prediction is a feature of language comprehension in general, and not a feature of comprehension limited to any specific language, it is possible for predictive processes to manifest in similar ways across languages within an individual. If response dominance patterns reflect a processing trait of the individual, then similar response dominance patterns should be observed in the learner's L1 and in their L2. That is, whether a learner bases their linguistic predictions on semantic information or syntactic information in L1 should predict whether that learner bases their L2 predictions on semantic or syntactic information as well. Individuals who show more positive RDI values (more P600-dominant responses) to L1 grammatical errors are thus expected to also show more positive RDI values to L2 grammatical errors, compared to individuals who show more negative RDI values (more N400-dominant responses) to grammatical errors in L1. However, if linguistic prediction somehow proceeds in very different ways across languages, then this relationship may not hold. For example, it is possible that speakers of languages with very strict word order rules, such as English, may have more individuals who rely on syntactic prediction for comprehension than speakers of languages with much more flexible word order rules, such as Russian. Nevertheless, all languages encode features or form and features of meaning, so both types of linguistic information can inform predictions in any given language. Even if there are fewer P600-dominant individuals in a flexible word order language, these individuals will still be expected to show more positive RDI values in their L2.

Furthermore, in order to develop predictions (in any domain), individuals primarily need access to reliable input, so the success of predictive mechanisms need not be limited by existing knowledge, although it may be facilitated by it. Therefore, not only are response dominance patterns expected to relate across grammatical rules that exist in both L1 and L2 (Bice & Kroll, 2017; Finestrat et al., 2020), but also for grammatical rules that are unique to the L2 and do not exist in the L1. Language and rule similarity have long been considered barriers to native-like L2 acquisition, but if individuals are relying on predictions rather than on L1 syntactic knowledge for acquisition, this obstacle may be mitigated. Moreover, individuals who build stronger syntactic predictions in L1 may also overcome this obstacle more easily than individuals who build stronger semantic predictions when it comes to the acquisition of unique L2 rules. In fact, individuals who showed larger P600 magnitudes to L1 syntactic violations were indeed found to acquire syntactic rules in an artificial language more quickly than learners who showed larger N400 magnitudes (Qi et al., 2017). It therefore appears that learners who attend more to syntactic information are faster at extracting L2 syntactic rules, deploying syntactic predictions, and adjusting those predictions given subsequent prediction errors than learners who attend more to semantic information. Thus, learners who have higher proficiency in a unique L2 rule are predicted to be individuals who attend more to syntactic cues in their L1.

One remaining consideration is that it is not immediately clear what the best outcome variable for measuring proficiency in a rule is. Accuracy, on the one hand, can reveal how sensitive a learner is to correct and incorrect rule use, but accuracy is an offline measure of comprehension informed by many other cognitive processes, such as memory, decision making, analogy, etc. This is one reason why ERPs have been such a powerful tool in language acquisition research, as many studies have found neural sensitivity to grammatical errors during online sentence comprehension even when learner's offline grammaticality judgment accuracy was at chance (McLaughlin et al., 2010; Osterhout et al., 2006, 2008). However, recall that since there is variability in ERP processing even among highly proficient native speakers, neither ERP component is a perfect measure of "end state" L2 acquisition. Calling back to data from Mclaughlin et al. (2010), is therefore unclear whether individuals at intermediate stages of learning show N400 effects to grammatical errors because they are still transitioning to a P600 or because they have already reached their ideal processing pattern. One way to tease this apart is to control for L2 language experience and L2 proficiency, two variables that have often been shown to influence development of neural sensitivity to grammar (Kotz, 2009). Learners with more L2 experience are typically more sensitive to errors, but when learners with the same level of L2 experience continue to show differences in L2 proficiency, this can be thought to reflect L2 learning aptitude (Birdsong, 2006; Kaan, 2014; Rossi et al., 2006a; C. M. Weber-Fox & Neville, 1996). If patterns of L1 neural processing still predict L2 rule accuracy and patterns of L2 neural processing and over and above these effects, then it can be argued that learners' linguistic predictions influence L2 acquisition over and above the individual's aptitude for learning an L2.

1.4.2 *The Syntactic Sensitivity account*

The Syntactic Sensitivity account of individual differences in ERP grammar processing proposes that individuals vary with regard to their sensitivity to grammatical form, such that P600-dominant individuals have increased language analytic ability compared to N400-dominant individuals. Language analytic ability refers the conscious awareness of the roles that words play in language and the capacity to detect, make inferences about, and generalize linguistic rules. Under this proposal, variability in the capacity to detect violations of linguistic rules would primarily drive differences in L1 processing, while differences in L2 learning would rely on all facets of this skill. This hypothesis stems from the traditional interpretation of the N400 as an index of semantic analysis and integration and the P600 as an index of syntactic analysis and repair. Not only do syntactic errors typically reveal P600 effects in groups of speakers, but P600 effect magnitudes also tend to be larger in syntactically complex or ambiguous structures (Carreiras et al., 2004; Frisch et al., 2002; Gouvea et al., 2010; Kaan et al., 2000; Kaan & Swaab, 2003; Mehravari et al., 2015; Phillips et al., 2005). Individuals who show larger P600 magnitudes in response to syntactic violations may therefore engage grammar processing streams to a greater degree than N400-dominant individuals.

1.4.2.1 Language analytic ability in L2 learning

Although a strong case has thus far been made for predictive roles for the N400 and P600 components, this prediction interpretation needs to contend with the large body of research that has long implicated language analytic ability in efficient L2 learning and processing (Dörnyei & Ryan, 2015; Dörnyei & Skehan, 2003; Naiman, 1996; Robinson, 2002). Language analytic ability is more commonly referred to in the L2 literature as grammar aptitude and is assessed in numerous language aptitude measures, including the Modern Language Aptitude Test (MLAT; Carroll & Sapon, 1959), the LLAMA Language Aptitude Test (Meara, 2005), and the Pimsleur Language Aptitude Battery (PLAB; Pimsleur et al., 2004). These assessments typically test for several specific linguistic skills believed to be integral for language learning success (Carroll & Sapon, 1959), and have been fairly successful at predicting an individual's L2 learning potential. However, in a meta-analysis of 33 studies investigating the role of language aptitude in L2 grammar acquisition, Li (2015) demonstrated that L2 grammar proficiency is specifically related to language analytic ability over other aptitude features, such as phonetic coding ability or vocabulary learning. Language analytic ability is especially central for language learning in older learners relative to younger learners (Abrahamsson & Hyltenstam, 2008; DeKeyser, 2000; DeKeyser et al., 2010; Granena & Long, 2013; Harley & Hart, 1997). For instance, scores on the MLAT Paired Associates task, a measure of vocabulary aptitude, were positively correlated with several proficiency measures in a group of students who began French immersion in the first grade (Harley & Hart, 1997). Students who began immersion in seventh grade, on the other hand, instead showed a positive correlation between MLAT Words in Sentences, a measure of grammar aptitude, and those same proficiency measures. It has even been suggested that a high degree of language analytic ability is what allows adult learners to eventually pass for native speakers in everyday interactions (Abrahamsson & Hyltenstam, 2008).

Language analytic ability has been shown to predict L2 acquisition in ERP research, as well. L2 Spanish learners' P600 magnitudes and d-prime scores (a Signal Detection Theory measure of accuracy) in response to noun-adjective agreement errors were correlated with overall performance on the MLAT, but this relationship was primarily driven by performance on the MLAT Words in Sentences subcomponent (Bond et al., 2011). Given this relationship, it would appear that L2 French learners in McLaughlin et al.'s (2010) multi-session study who demonstrated P600 effects to grammatical violations in the middle session had stronger language analytic skills than those learners who were still showing N400 effects. This might be one reason that N400 magnitudes during this session were not correlated with proficiency in the grammatical rule but P600 magnitudes were. If it is assumed that the N400-group was comprised of slower learners who had poor grammar aptitude, it is possible that it was fairly uniform with regard to power language analytic skill and proficiency. The P600-group, on the other hand, was comprised of better learners for whom the development of a P600 response was driven by their increased grammar learning aptitude. Thus, P600 magnitude was tied not just to proficiency but also to language analytic ability. Further, language analytic ability may be related not only to the faster acquisition of an L2 rule, but also its slower attrition. The transition from an N400 response to a P600 response over the course of learning has been found to reverse as time since instruction increased (Osterhout et al., 2019). English learners of Finnish showed a decrease in P600 magnitudes and an increase in N400 magnitudes in response to an L2-unique phonological rule violation over a period of attrition. However, variability in P600 magnitude was evident even among learners who had similar durations since the last day of instruction, suggesting that aptitude may have been responsible for helping some learners maintain sensitivity to the L2 rule.

1.4.2.2 Language analytic ability in L1 processing

Much less work has considered the role of language analytic ability in explaining variability in L1 comprehension. This is due in part to the long-held belief that there are far fewer differences in language proficiency among native speakers of a language than its L2 learners. There is also an assumption that language analytic ability is not as relevant for L1 development as it is for L2 acquisition because children rely on implicit learning mechanisms compared to adult learners (Bley-Vroman, 1989; Dąbrowska, 2018; DeKeyser, 2000, 2008). For instance, Li's (2015) meta-analysis found that the relationship between overall aptitude or language analytic ability and L2 grammar proficiency was moderated by learning context, where the strength of the relationship was improved in explicit compared to implicit learning contexts. Nevertheless, there is evidence that L1 speakers do in fact vary in proficiency, as already demonstrated herein, as well as both direct and indirect evidence that L1 acquisition and processing is informed by language analytic ability.

First, the grammar section of the MLAT specifically tests a learner's language analytic ability in the native language. Questions are constructed such that a model sentence is presented alongside a test sentence. The learner's task is to identify the word in the test sentence that plays the same role as a specified word in the model sentence. In this way, participants are asked to turn their language analytic ability onto their own native language. It is evident that native speakers are not homogeneous in their analytic ability or, relatedly, their L1 proficiency because native speakers show a good deal of variability in these scores. Furthermore, given that the MLAT is considered the "gold standard" of aptitude assessments that has been reliably predicting L2 learning success for over 50 years, it is notable that the best measure for predicting L2 grammar acquisition is based entirely in L1 knowledge. This stands in contrast to the grammar portions of the LLAMA and PLAB aptitude test, which assess a learner's ability to detect, infer, and generalize rules in an artificial language.

Moreover, there is new data that this kind of grammatical inferencing test of analytic ability does explain variability in L1 grammar use and knowledge (Dąbrowska, 2018, 2019). For instance, Dąbrowska (2018) found that the PLAB grammar aptitude measure explained 10% of the variability in grammatical comprehension of a variety of simple and complex English grammatical constructions in a group of native English speakers. Not only do differences in analytic ability explain performance in adulthood, but there is evidence that this ability emerges early in childhood and can influence children's development as they acquire reading and writing skills in school. Language performance tested in children at 5-years-old was found to predict language analytical abilities in those same children 10 years later (Dörnyei & Ryan, 2015; Skehan & Ducroquet, 1988; Wells, 1985). Furthermore, high schooler's L2 proficiency was showed to be predicted by assessments of those students' reading skills as early as the fourth grade (Sparks et al., 2008, 2009). If early L1 performance has such a lasting and predictive effect on L1 analytic abilities and L2 acquisition later in life, it is hard to imagine that it doesn't already interact with language development at its earlier stages and contribute to the variability in grammar processing observed in adulthood.

Finally, L1 ERP dominance patterns have recently been directly implicated in adult artificial language learning (Qi et al., 2017). In the first and only published study of its kind, English speakers' ERPs were recorded in response to semantic and syntactic violations in English and participants were then trained and tested in an artificial language over the course of 3 sessions. Participants were categorized as N400-dominant or P600-dominant in response to both semantic and syntactic violations in English and their vocabulary, semantic, and syntactic proficiency in the artificial language was compared to their L1 processing. Qi and colleagues interpreted dominance as a proxy measure of aptitude, such that P600-dominant individuals in the syntax condition had better grammar aptitude and N400-dominant individuals in the semantic condition had better vocabulary and semantic aptitude. Data revealed that individuals who were N400-dominant in response to semantic errors did in fact show better vocabulary and semantic proficiency in the artificial language, while individuals who were P600-dominant in response to syntactic errors showed better syntactic proficiency. Both patterns were observed as early as the first day of training and testing. These data suggest that P600 dominance is related to syntactic aptitude and that L1 processing patterns can influence L2 acquisition. Given that L1 grammar processing variability and language analytic ability have both been shown to predict L2 grammar proficiency, it is thus possible that L1 grammar processing and aptitude are related to one another, such that individuals who are P600 dominant also have better grammar aptitude.

1.4.2.3 Implications for individual differences

To reiterate, the Syntactic Sensitivity account of individual differences in grammar processing proposes that individuals vary with regard to their sensitivity to grammatical form. The evidence presented thus far shows that language analytic ability and L1 processing variability, specifically P600 effect magnitude, have both been implicated in explaining differences in L2 grammar acquisition and have recently begun to be related to each other. If the P600 is a measure of sensitivity to violations of grammatical form, then individuals who are P600-dominant are expected to have increased language analytic ability compared to N400-dominant individuals. N400-dominant individuals, on the other hand, may be engaging in “good-enough” processing (Ferreira, 2003; Ferreira et al., 2002; Ferreira & Patson, 2007). The Good-Enough view of language comprehension proposes that readers may create a representation of a sentence that is sufficient for getting the overall meaning being conveyed but does not delineate the exact relationship between the elements of that representation. Ferreira (2003) showed the comprehenders incorrectly identified the agent in passive sentences such as “*The customer was thanked by the clerk*” about 20% of the time compared to the active sentence “*The clerk thanked the customer*.” Performance was even worse for more complex syntactic structures, e.g., “*It was the customer that was thanked by the clerk*”. Readers’ comprehension was therefore good enough: it represented the gist of the meaning and context, but the fidelity of the event representation was lacking. This kind of shallow grammatical processing has been implicated in the interpretation of the N400 (Ferreira & Patson, 2007; Mehravari et al., 2017; Rabovsky et al., 2016). It is possible then that N400-dominant individuals engage with sentences containing grammatical errors in a good-enough manner. Again, although "*The roses *<u>grows</u> in the garden*" is ungrammatical, the gist of the sentence is largely unchanged due to the error and there is no need to recruit deeper processing for syntactic re-analysis. The N400 may instead reflect noticing of the discrepancy in plurality, or, even more simply, a difference in expected word length.

1.4.2.3.1 Predictions for individual differences in L1 processing

If the P600 effect is characterized as a measure of sensitivity to grammatical form, then differences in individuals' language analytic ability will manifest as variability in ERP response dominance patterns. The more P600-dominant an individual is, the better that individual is expected to score on a measure of grammar aptitude. Given that more positive RDI values indicate more P600-dominant responses, RDI values to syntactic violations should be positively correlated with grammar aptitude scores. It is possible that the strength of this relationship will also be modulated by the type of syntactic anomaly the RDI value is computed from. For example, sentences containing phrase structure violations may show the strongest correlation with grammar aptitude. In this type of error, words appear in an order that is prohibited by a language's grammar (e.g., "*Their niece ate soup *<u>her</u> without spilling a* drop") and thus cannot be easily repaired without deep syntactic restructuring. However, agreement violations may show weaker relationships with language analytic ability because they occur on the word form and do not disturb the underlying structure in the same way that word order violations do. Agreement requires a correspondence between syntactic and semantic information for correct morphosyntax in verb tense agreement conditions (e.g., "*Mandy was cooking /*cook dinner for her visitors*") or verb form selection in subject-verb agreement conditions (e.g., "*Later tonight, the committee has/*have an important meeting*"). In both conditions, semantic cues can contribute to the initial interpretation and subsequent re-interpretation that would tax grammar processing streams to a lesser degree. In the case of verb tense, for example, the meaning can be reinterpreted as occurring in the future tense without adjusting the syntactic structure if the comprehender chooses to represent *cook* as a grammatical verb form and use it to override the semantics of the past tense *was*. The influence of semantics on reinterpretation could be even more apparent in the subject-verb agreement condition when a collective noun such as *committee* is used. Collective nouns (e.g., *committee, band, jury*) are themselves semantically plural, being composed of many entities, but take syntactically singular verb forms. A comprehender could reinterpret a collective noun agreement violation without having to re-analyze the sentence with a new verb form by simply tapping into the semantic plurality of the noun, thus rendering the verb grammatical. In short, language analytic ability may be most strongly correlated with word-order violations, followed by verb-tense violations, and finally by collective noun-verb agreement violations.

With regard to whether ERP response dominance patterns reflect a trait of the individual, both the Top-Down Prediction account and the Syntactic Sensitivity account would predict that RDI values should be correlated across syntactic conditions. Again, if an individual is more sensitive to grammatical form, then they should be more sensitive to syntactic violations and have a more P600-dominant response profile regardless of the type of anomaly a violation contains. Thus, more positive RDI values (more P600-dominant) across violations would be observed under this account as well. Contrary to the predictions of the Top-Down Prediction account, however, the Syntactic Sensitivity account would not predict a relationship between the way an individual processes syntactic violations and semantic violations. Semantic continuations may vary in expectancy, but they do not vary in form or grammatical agreement from one another or violate syntactic rules. For instance, the sentence "*Finally at home, Perry took out his keys to unlock the door/kennel/*traffic…*" differs only in the semantic fit of the noun ending. No grammatical rules are violated, and no syntactic cues would influence reinterpretation. If the P600 reflects sensitivity to grammatical form, then individuals who are typically more sensitive to syntactic violations should not respond in a systematically different way to semantic expectancy anomalies than individuals who are less sensitive to grammatical form. The presence or absence of a relationship between semantic and syntactic conditions is therefore the critical comparison that will disambiguate which of the two hypotheses better explains variability in L1 grammar processing.

Finally, the Syntactic Sensitivity account would not predict any relationships between RDI values and beta desynchronization. First, if the P600 is a measure of sensitivity to grammatical form and not an index of predictive processes, there should be no significant relationships between RDI values and beta desynchronization in any syntactic condition. Comprehenders who are more sensitive to grammatical form will recruit syntactic processing streams to a greater degree to deal with a syntactic violation compared to those who are less sensitive, but the violation itself should not be more or less surprising to either group. The same should be true for semantic violations. Again, if there are no violations to form in the semantic conditions, then semantic conditions should not be any more or less surprising to individuals with high analytic ability and more positive RDI values to syntactic violations that individuals with low analytic ability and more negative RDI values.

1.4.2.3.2 Predictions for individual differences in L2 learning

Variability in learners’ proficiency in L2 grammatical rules is expected to correlate with language analytic ability under the Syntactic Sensitivity account. This relationship should hold both for L2 rules that exist in the L1 as well as L2 rules that are unique to the L2. In fact, the correlation between sensitivity to grammatical form and L2 rule proficiency may be stronger for a unique L2 rule than a similar L2 rule. For a familiar rule, learners’ L1 knowledge may drive acquisition more so than grammar aptitude given that the learner is already practiced in its use. L1-L2 similarity may also result in learners reaching high proficiency in the familiar rule during earlier stages of acquisition. In this case, correlations with language analytic ability may be altogether absent. However, acquisition of the unique L2 rule cannot be informed by L1 knowledge to the same degree as a similar L2 rule, so grammar aptitude would play a driving role in developing its proficiency. Unique L2 rules are also more difficult to acquire and take more time to master. In fact, English learners of German continued to struggle with grammatical gender agreement, a rule that does not exist in English, even with an average 20 years of experience and German residency (Hopp, 2013). Thus, after controlling for years of L2 experience, learners who gain higher proficiency in the unique L2 rule may be able to do so thanks to increased language analytic ability. Again, higher proficiency would be assessed not just by the magnitude of learners’ ERP responses to errors, but by their accuracy to grammaticality judgments as well.

If the P600 is an index of sensitivity to grammatical form, then a learner’s sensitivity to grammatical form in L1 should transfer to similar L2 rules but not to unique L2 rules. More positive RDI values, which reflect a greater degree of sensitivity to grammatical form, should therefore be correlated within a familiar rule across languages. Again, given that the rule is expressed similarly in both languages and learners already have some level of sensitivity to its form in L1, neural responses in L2 are expected to follow suit. However, it would be far more difficult for learners to acquire sensitivity to grammatical forms that are unfamiliar to them because they would be unable to rely on L1 knowledge. On the other hand, experience with the L2 might be able to help inform sensitivity to unique L2 rules. As learners gain L2 experience, they become more familiar with the demands of the language and are able to contextualize the purpose of unique L2 rules given its constraints. For example, if a learner's L1 relies on strict word order over morphology for delineating relationships between elements of an event, acquiring morphological rules in a language with flexible word order will be quite challenging. The learner is unfamiliar with how to use morphology for building relations in L1 or why it is important. However, as experience with the L2 increases and the learner becomes aware of the relevance of morphology for conveying the appropriate correspondence between actors in a sentence, sensitivity to morphological rules will increase. Thus, L2 experience will drive sensitivity to grammatical form in unique L2 rules more so than L1 syntactic sensitivity. L2 experience would include years of formal instruction and L2 grammar proficiency, but also experience processing other L2 rules. Thus, contrary to the predictions of the Top-Down Prediction account, L1 processing patterns are not predicted to influence L2 unique rule learning, particularly not above L2 similar rule processing patterns or measures of L2 experience. The presence or absence of a correlation between L2 unique rule proficiency and L1 ERP dominance is therefore the critical comparison that will disambiguate which of the two hypotheses better explains variability in L2 grammar learning.

Lastly, the Syntactic Sensitivity account would not predict any relationships between L2 unique rule proficiency and beta desynchronization. As in the case of L1 processing, learners who are more sensitive to grammatical form are expected to recruit syntactic processing streams to a greater degree to deal with that violation, but the violation itself should not be more or less surprising. Moreover, whether an L2 learner relies more on syntactic processing streams to deal with an L2 unique rule or not, this variability should not be influenced by how surprising a learner finds a violation to be in L1, especially given that the learner does experience surprisal to the unique rule in L1. Thus, there should be no relationship between L2 unique rule proficiency and L1 beta desynchronization, particularly not above L2 experience.

1.5 Overview and goals of the present studies

Two experiments were used to adjudicate between the two hypotheses described above regarding individual differences in grammar processing. In Experiment 1, native monolingual English speakers read sentences containing three different syntactic violations as well as sentences containing continuations that varied in degree of semantic expectancy. In Experiment 2, native Chinese speakers of English read Chinese sentences containing a grammatical violation and English sentences containing a similar syntactic error as well as sentences containing violations of a rule that does not exist in Chinese. In both experiments, EEG and grammaticality judgments were recorded and the EEG signal was analyzed to measure ERP effect magnitudes and degree of beta desynchronization. Three primary analyses were conducted to test the validity of each hypothesis by attempting to explain variability in the base condition of each experiment using a measure of language analytic ability or neural measures of processing in a comparison condition. The base condition is a condition where individual differences in grammatical processing have previously been observed. In Experiment 1, the base conditions are the sentences containing syntactic violations, whereas in Experiment 2, the base condition is the unique L2 rule. The comparison condition is a condition that would be predicted to be related to the base condition under the Top-Down Prediction account but not under the Syntactic Sensitivity account. In Experiment 1, the comparison conditions are the sentences containing semantic violations, whereas in Experiment 2, the comparison condition is the L1 rule violation. The three analyses are briefly described here.

1.5.1 *Is variability in grammar related to variability in language analytic ability?*

To test the predictions of the Syntactic Sensitivity account that variability in grammatical processing is related to variability in language analytic ability, RDI values in the base condition will be related to scores on a grammar aptitude test. If such correlations exist, the data would provide evidence in support of the Syntactic Sensitivity account and against the Top-Down Prediction account.

1.5.2 *Does variability in grammar processing reflect variability in the nature of individual's linguistic predictions in language processing?*

To test the prediction of the Top-Down Prediction account that variability in grammatical processing is related to variability in the nature of individual’s linguistic predictions, RDI values and beta desynchronization in the base condition will be related to RDI values and beta desynchronization in the comparison condition. Recall that RDI values and beta desynchronization index to separate but complementary features of prediction. RDI values describe the *type* of linguistic prediction being made whereas changes in beta power reveal the *strength* of linguistic prediction. If type or strength of predictions are related across the base and comparison condition, this would provide evidence in support of the Top-Down Prediction account and against the Syntactic Sensitivity account.

1.5.3 *Does variability in type of linguistic predictions relate to variability in strength of*

linguistic predictions?

To further test the prediction of the Top-Down Prediction account, RDI values in the base condition will be related to Beta values in the base condition as well as the comparison condition. Whereas the aim of the second analysis is to compare a single feature of prediction (type or strength) across conditions, the aim of this third analysis is to compare features of prediction against each other. Thus, the goal is to assess whether individual differences in the type of predictions individuals make can be explained by variability in the strength of predictions individuals make. In other words, is it possible that using syntactic or semantic information for linguistic prediction is better for building stronger predictions? If type of prediction and strength of prediction are related to one another, this will provide further evidence in support of the Top-Down Prediction account and against the Syntactic Sensitivity account.

Chapter 2. METHODS

2.1 Experiment 1

2.1.1 *Participants*

Eighty-three healthy adults aged 18-37 (M = 19.95, SD = 3.14; 51 female) were recruited for participation from the University of Washington Psychology Subject Pool. All participants were screened as monolingual native English speakers born in the United States who began acquiring English at birth and were not exposed to any languages besides English before the age of 10. Socio-demographic variables relating to age, years of education, socio-economic status, language background and fluency in English and other languages, and handedness were collected for all participants. All participants provided informed consent according to the standards outlined by the University of Washington Institutional Review Board and were compensated for their time with course credit. Two participants reported being exposed to Tagalog before the age of five and were consequently excluded from analyses. Data from 57 participants (Age: M = 19.98, SD = 2.87; 35 female) who completed both EEG data collection sessions with no more than 25% trial rejection rate per session (**see Analyses**) was used for subsequent analyses.

2.1.2 *Materials*

Participants were presented with two blocks of sentences: the syntax block and the semantics block. The order of the blocks was counterbalanced across subjects and sessions.

2.1.2.1 Syntax block stimuli

The syntax block was composed of sentences containing three different kinds of violations: word order violations, verb tense violations, and subject-verb agreement violations (see Table 1). The word order condition contains a violation of syntactic structure where reanalysis requires insertion or movement of branches in deep syntax (e.g., *Their niece ate* <u>*her*</u> *soup/soup ** <u>*her*</u> *without spilling a drop*). The verb tense and subject-verb agreement conditions contain violations of verb form. In the verb tense condition, the form of the verb is missing the appropriate morphological ending *-ing* (e.g., *Mandy was* <u>*cooking*</u>*/** <u>*cook*</u> *dinner for her visitors*), whereas the entire form of the verb is inappropriate in the subject-verb condition (e.g., *Later tonight, the committee* <u>*has*</u>*/** <u>*have*</u> *an important meeting*). Appropriate verb form in both conditions requires a coordination of semantic and syntactic information with the word that occurs before it. In the verb tense condition, the preceding auxiliary (e.g., *is*/*are*/*was*/*were*) signals (among other things) the progressive tense, activating the morpho-syntactic rule that the following main verb must end in *-ing*. Although tense requires a coordination of structure, it also provides semantic information as to whether the action was ongoing or completed. In the subject-verb agreement, the verb selected for the subject must also agree with the plurality of the subject, a semantic feature of the noun (e.g., *The* <u>*cat was*</u> *eating* vs. *The* <u>*cats were*</u> *eating*). However, in order to put semantic and syntactic cues at greater odds with each other, collective nouns were used as subjects in this condition, which will henceforth be referred to as the collective noun condition (e.g., *Later tonight, the committee* <u>*has*</u>*/** <u>*have*</u> *an important meeting*). Collective nouns are nouns such as *committee* or *team* that give a single name to a collection of entities. As a result, collective nouns are syntactically singular, requiring singular verb forms (*has* vs. *have*), but semantically plural. Therefore, the syntactic error in this condition could be reanalyzed either by applying to correct verb form or by accessing the semantic plurality of the collective noun.

Table 1. Example Sentences in Syntax Block. The critical word is underlined.

Condition	Example
Word Order	*Their niece ate <u>her</u> soup/soup *<u>her</u> without spilling a drop.*
Verb Tense	*Mandy was <u>cooking</u>/*<u>cook</u> dinner for her visitors.*
Collective Noun	*Later tonight, the committee <u>has</u>/*<u>have</u> an important meeting.*

The word order condition contained 50 items and sentences were produced using two different types of phrase-structure violations within the verb phrase to increase variety of violations in the stimuli. Half of the sentences (25) were constructed using modal verb + main verb + determiner + noun verb phrases (e.g., *The chef will chop <u>these</u> onions for the sauce*). Violations were produced by moving the determiner + noun after the modal verb (e.g., *The chef will *<u>these</u> onions chop for the sauce*). Sentence length ranged from 8 to 10 words (M = 9.44, SD = 0.85).

The second half of the word order condition (25) was composed of sentences containing verb + pronoun + noun verb phrases (e.g., *Their niece ate her <u>soup</u> without spilling a drop*). Violations were constructed by reversing the order of the pronoun and noun (e.g., *Their niece ate soup *<u>her</u> without spilling a drop*). To ensure that the violation occurred only at the point when the noun was ungrammatically followed by a pronoun, sentences were written using plural nouns or mass nouns (nouns that refer to something that cannot be counted, e.g., *hair, water*, *oxygen*) such that they remained grammatical until the point of the pronoun. Sentence length ranged from 7 to 11 words (M = 8.64, SD = 0.97).

Critical words in the grammatical word order conditions appeared in the range of the 3rd to 6th position in the sentence (M = 4.52, SD = 0.75) and critical words in the ungrammatical word order conditions appeared in the range of the 3rd to 7th position in the sentence (M = 4.52, SD = 1.00). Although the position of the critical word shifted between conditions within each word order type, there was no statistically significant difference in the average position of the critical word between the grammatical and ungrammatical versions of the word order conditions, t(48) = 0, p > .05, minimizing limitations in condition comparisons.

Fifty sentences were constructed for the verb tense condition. Verb phrases in the verb tense condition were constructed with an auxiliary verb + main verb structure (e.g., *Mandy was <u>cooking</u> dinner for her visitors*). Violations were produced by removing the -*ing* morpheme from the main verb (e.g., *Mandy was *<u>cook</u> dinner for her visitors*). Sentence length ranged from 7 to 11 words (M = 9.16, SD = 0.99).

Sentences in the collective noun condition were constructed using 50 collective nouns followed either by the verb *to be* or *to have* such that singular vs. plural verb agreement was overt in the verb form (e.g., *Later tonight, the committee <u>has</u> an important meeting*). Violations were produced by changing the plurality of the verb from singular to plural (e.g., *Later tonight, the committee *<u>have</u> an important meeting*). Given that verb agreement with collective nouns may vary over time (e.g., agreement with *police* was singular in American English until ca. 1850 according to a Google NGram search; Michel et al., 2011) and across dialects (e.g., British English allows for plural agreement with collective nouns, *The team are starting the presentation soon*), a sentence completion task was conducted with 37 undergraduate students at the University of Washington to ensure that participants found the chosen collective nouns to be syntactically singular. All participants were monolingual native English speakers born in the United States who began acquiring English at birth and were not exposed to any languages besides English before the age of 10. Participants provided informed consent and were compensated for their time with course credit. Two experimental lists were constructed containing 25 collective noun sentences and 25 sentences with plural count nouns (e.g., *pencils, restaurants, lights*) where the verb was replaced with a blank (e.g., *Later tonight, the committee* _____ *an important meeting*). Participants were instructed to fill in the blank with the verbs *to be* or *to have* and ensure that the sentence was grammatical. Sentences using *to have* were always written in the present tense to avoid use of *had,* which is not overtly discriminable between singular and plural agreement. On average, collective noun sentences were completed with a singular verb form 96% of the time (SD = 0.04, Range: 0.85-1.00). One item containing the collective noun *majority* was removed due to a low proportion of singular sentence completions (51.35%) and replaced with an item containing the collective noun *bundle*. The relative frequency of *bundle* + *is/was/has* phrases compared to *bundle* + *are/were/have* phrases in the Corpus of Contemporary American English (COCA) (Davies, 2008) was 0.91 and 0.89 in a Google NGram search of the American English corpus between the years 2000-2008 (Michel et al., 2011). Sentence length ranged from 7 to 12 words (M = 9.86, SD = 1.13).

Two experimental lists of 150 sentences were constructed for the syntax block containing 50 sentences from the word order, verb-agreement, and collective noun conditions each. Half of the sentences in each condition were grammatical and the other half were ungrammatical such that there were 25 items per each syntax by grammaticality condition. Each pseudorandomized list contained only one version of each experimental sentence such that no participant saw any given item in both grammatical and ungrammatical conditions. Participants were pseudorandomly assigned to one of the stimulus lists.

2.1.2.2 Semantic block stimuli

The semantics block contained sentences whose continuations were either semantically expected, unexpected but plausible, or implausible (e.g., *Finally at home, Perry took out his keys* *to unlock the* *<u>door</u>/<u>kennel</u>/?<u>traffic</u> for Stacy*). A total of 90 high constraint sentences were constructed such that their noun continuations were highly expected and piloted through a series of cloze studies to ensure cloze probabilities of expected endings were high. Participants were recruited from the University of Washington (n = 89) and Amazon Mechanical Turk (n = 206) and were all monolingual native English speakers born in the United States who began acquiring English at birth and were not exposed to any languages besides English before the age of 10. Participants at the University of Washington were compensated with course credit and participants on Amazon Mechanical Turk were paid for their time. Participants were presented with the sentences up to the point of the critical word, which was replaced with a blank, and were instructed to complete the sentence with three possible one-word continuations. Each item in the final set of stimuli was rated by either 89, 115, or 110 participants and cloze probability was computed as the percent of time a given word appeared as the first provided continuation. Expected continuations for the final set of stimuli had a mean cloze probability of 0.84 (SD = 0.08, Range: 0.69-0.99). Unexpected continuations had a mean cloze probability of 0.001, calculated as the percent of time the unexpected noun appeared as any of the three provided continuations (SD = 0.003, Range: 0.000-0.009). Continuations that made the sentence semantically ill-formed were chosen for the implausible condition and were checked against data from all of the cloze studies to ensure they never appeared as a continuation.

To ensure that the continuations were indeed expected, unexpected, and implausible given the continuation, Latent Semantic Analysis (LSA) values were computed to measure the strength of the semantic relationship between each continuation and its sentential context for each condition (Landauer et al., 1998; *Latent Semantic Analysis @ CU Boulder*, 1998). Following the expected pattern, a Tukey HSD test showed that expected continuations (M =

0.32, SD = 0.13) had higher LSA values than unexpected continuations (M = 0.22, SD = 0.11), *p* < 0.001, and unexpected continuations had higher LSA values than implausible continuations (M = 0.16, SD = 0.10), *p* = 0.001. Continuations were also controlled for lexical features that might affect the magnitude of the N400. A one-way ANOVA showed there were no significant differences in the frequency of continuations across the three conditions according to COCA, *F*(2, 267) = 1.55, *p* = 0.213. There was also no significant difference in word length across the three conditions, *F*(2,267) = 1.11, *p* = 0.331. Sentence length ranged from 8 to 16 words (M = 12.10, SD = 1.75), and the average critical word position was 8.73 (SD = 2.15).

Three experimental lists of 120 sentence were constructed for the semantics block containing 30 expected, 30 unexpected, and 30 implausible continuation sentences, and 30 filler sentences. Filler sentences contained syntactic violations of determiner-noun agreement in order to balance the number of well-formed items and violations in each experimental list and to vary the types of violations in the block (e.g., *It turns out that Patricia sat on that* <u>*package*</u>/**<u>packages</u>* *by mistake*). Determiner-noun agreement violations were chosen so that the critical word was also a noun. Half of the fillers violated singular determiner-noun agreement and the other half violated plural determiner-noun agreement (e.g., *Next Thursday, we will harvest these* <u>*carrots*</u>/**<u>carrot</u>* *for the market*). Each pseudorandomized list contained only one version of each experimental sentence such that no participant saw any given item in more than one of the three experimental conditions. Participants were pseudorandomly assigned to one of the three stimulus lists.

2.1.2.3 Behavioral tasks

2.1.2.3.1 Language Analytic Ability

Language analytic ability was assessed using the LLAMA F, the grammatical inferencing subcomponent of the LLAMA Language Aptitude Test as a measure of language analytic ability (Meara, 2005). The LLAMA is a computerized language aptitude test that assesses language-learning relevant skills in an L1-independent manner. All subcomponents test aptitude for learning an artificial or a largely unknown indigenous language. Previous work has shown no difference in overall or by-section performance between native speakers of English, Spanish, and Chinese (Granena, 2013), and or between native speakers of languages with different orthographic scripts, such as English, Arabic, and Chinese (Rogers et al., 2016, 2017). Participants completed the entire LLAMA battery but only LLAMA F scores were applicable for measuring grammar aptitude. The LLAMA F measures the ability to extract grammatical rules of an artificial language through explicit inductive learning, a critical factor in of Carroll's model of language learning aptitude (Carroll, 1965; Dörnyei & Ryan, 2015). The artificial language contains number and gender agreement, strict word order, varied preposition use, and phonetic feature agreement (i.e., voicing). In the learning phase, participants were presented with a set of 20 images that were each accompanied by a sentence describing the content of the picture in the artificial language. The sentence-picture pairs are composed in such a way as to create minimal pairs for the features of interest, but participants were not given any information about these features. Participants were instructed to learn as much as they can about the language by exploring the sentence-picture pairs over the course of five minutes. In the following test phase, participants were presented with 20 novel images, each accompanied by a grammatically correct and a grammatically incorrect sentence. Participants were instructed to select the sentence that correctly describes the picture in the artificial language and were provided feedback after each item. The test phase was not timed, and participants could earn a maximum of 100 points.

2.1.2.3.2 Working Memory

Participants completed the Advanced Symmetry Span Task, a visual complex working memory task, as a measure of language-independent working memory capacity (Draheim et al., 2018). In this task, participants were presented with a sequence of red squares presented one at a time in random locations of cells in a 4x4 matrix. The sequence was to be memorized and recreated after all items in the sequence were presented. To interrupt rehearsal, the presentation of each red square in the sequence was followed by a visual processing task that required participants to judge whether a figure presented within an 8x8 grid was symmetrical along the vertical axis. At the end of the set, participants recalled the sequence of red square locations in the order in which they were presented by clicking on the cells of an empty matrix. The length of the sets spanned from two to seven items and each set size was presented three times in a randomized order. Participants scored one point for each correctly recalled square location but only in sets that were recalled correctly in their entirety. The maximum number of points a participant could earn was 81.

2.1.2.3.3 English Proficiency Test

In order to assess participant’s language proficiency, participants completed an online test of English proficiency developed by a language teaching service called Transparent Language (*English Proficiency Test*, n.d.). The proficiency test contained four sections of multiple-choice questions that tested grammar, vocabulary, and reading comprehension skills. Grammar skills were tested in two ways. The first grammar section instructed participants to select the grammatically appropriate word to fill in the blank of a sentence and the second section instructed participants to select which of four underlined words in a sentence was grammatically incorrect. Grammar proficiency scores were calculated by averaging scores from the two grammar sections. The test was not timed, and participants could earn up to 100 points. This measure was added after data collection had begun, so proficiency test scores were only available for 38 participants.

2.1.2.3.4 Reading Experience

Reading experience was assessed as a measure of language experience using the updated version of the Author Recognition Test (ART; Acheson et al., 2008; Stanovich & West, 1989), which has been shown to be a reliable test of reading skill and experience (Mol & Bus, 2011; Moore & Gordon, 2015). Participants were given a list of 130 names, half of which were names of real fiction and non-fiction authors and half of which were fabricated. Participants were told that some of the author names were real and some were not and instructed to mark the names of authors they were confident were real authors as incorrect choices would incur a penalty. Reading score was computed by calculating the difference between total hits and total false alarms, with a maximum possible score of 65. This measure was added after data collection had begun, so ART scores were only available for 41 participants.

2.1.3 *Procedures*

Participants were brought into the lab for two sessions to complete the syntax and the semantics blocks. The order of the sessions was counterbalanced across subjects. At the beginning of each session, participants completed questionnaires, the English Proficiency Test, and the ART while experimenters applied the EEG caps and electrode arrays. Participants were then brought into the data collection room where they were seated in a comfortable chair in a quiet, dimly lit room across from a CRT computer monitor. Participants were instructed to relax and sit comfortably while they read the sentences to themselves. They were informed that each sentence would be followed by a YES/NO prompt that would cue them to make a response as to whether the sentence was a good sentence in English, meaning that the sentence was grammatical and that it made sense. Participants were given a mouse and instructed to use the two buttons to make their judgments. The response buttons were counterbalanced across participants who were pseudorandomly assigned to use either the left or right mouse button for the “yes” response. In order to minimize artifacts, participants were encouraged to minimize blinking and movement during sentences, and to take as much time as they needed between sentences and sentence blocks to relax and rest their eyes before advancing.

All stimuli were presented with white letters in the center of a black background. Each trial began with a 400ms fixation cross (+) with a 100ms inter-stimulus interval and was followed by the sentence, which was presented one word at a time. Each word appeared on the screen for 400ms with a 200ms inter-stimulus interval. After the last word of the sentence disappeared from the screen, participants saw the YES/NO prompt and made their acceptability judgments. There was no time limit for providing a response, but participants were encouraged to go with their first instinct and not spend too much time deciding. Between each sentence, participants saw a READY? screen until they pressed a button to advance to the next sentence. Following the sentence-reading task in the first session, participants additionally completed the LLAMA Aptitude Test and Symmetry Span tasks.

2.1.4 *EEG recording and processing*

Continuous EEG was recorded from 32 Ag/AgCL scalp electrodes using the ActiveTwo Biosemi System (BioSemi, Amsterdam, The Netherlands) in accordance with the extended international 10/20 system (Nuwer et al., 1998). Electrode positions included four midline sites

(Fz, Cz, Pz, Oz), nine pairs of medial sites (left hemisphere: Fp1, AP3, F3, FC1, C3, CP1, P3, PO3, O1; right hemisphere: Fp2, AP4, F4, FC2, C4, CP2, P4, PO4, O2), and five pairs of lateral sites (left-hemisphere: F7, FC5, T7, CP5, P7; right hemisphere: F8, FC6, T8, CP6, P8). Eye blinks were recorded with an electrode placed under the left eye at the infraorbital ridge (vEOG, vertical electrooculogram) and horizontal eye movement were recorded with an electrode placed on the outer canthus of the right eye (hEOG, horizontal electrooculogram). Two additional electrodes were placed on the left and right mastoid to serve as reference. According to the BioSemi system design, a Common Mode Sense active electrode and a Driven Right Leg passive electrode were placed on either size of electrode POz, which served as the ground. The EEG signal was referenced online to the left mastoid, sampled at 256 Hz, and bandpass filtered between 0.16Hz and 100Hz.

All processing was carried out using the EEGLab (Delorme & Makeig, 2004) and ERPLab (Lopez-Calderon & Luck, 2014) analysis packages developed for the MATLAB environment. Independent Component Analysis (ICA) was performed on the data to remove artifacts related to blinks. The EEG data was decomposed into a set of underlying components, noisy components related to blinks were identified and rejected, and the data was reconstructed without the artifactual components. This artifact-corrected data was then used for subsequent preprocessing steps for ERP and time-locked frequency analyses.

2.1.4.1 ERP data processing

The ICA-ed EEG signal was bandpass filtered offline from 0.1 to 30 Hz, re-referenced to the average of the left and right mastoid, and segmented into individual epochs around the critical word in each condition from -100 ms to 1205 ms. Epochs were further marked for exclusion if they contained changes within a sliding 200ms window that were greater than 100 µV in Cz, vEOG, and hEOG electrodes. In addition, epochs were removed if they contained artifacts relating to excessive alpha amplitudes, electrode drift, or electrical drive in electrodes within a region of interest (ROI) corresponding to centro-parietal electrodes contained in the primary dependent variable of interest for individual differences analyses, the RDI. If an epoch showed voltage steps in excess of 65 µV within electrodes used to calculate the RDI (C3, Cz, C4, CP1, CP2, P3, Pz, P4) (Tanner, 2019; Tanner et al., 2014; Tanner & Van Hell, 2014), it was rejected.

Participants whose rejection rates were in excess of 25% for any experimental condition were excluded from analyses. All remaining subjects in Experiment 1 had an average rejection rate of 4% for the syntax block (SD = 0.03) and 4% for the semantics block (SD = 0.04). Only data from artifact-free trials that were judged correctly by participants were included in subsequent analyses. Epochs were baseline-corrected from -100 ms to 0 ms and averaged across trials within each condition for each participant.

2.1.4.2 Time-frequency data processing

The ICA-ed EEG signal was re-referenced to the average of the left and right mastoid and segmented into individual epochs around the critical word in each condition from -2 sec to 2 sec. For effective comparison of ERPs and time-locked frequency changes, any trials that were excluded from ERP analyses due to artifacts were also excluded from time-locked frequency analyses. Time-frequency analysis was used to quantify event-related spectral perturbations (ERSPs) (Delorme & Makeig, 2004). Mean ERSP was computed for ROI channels within each condition using morlet wavelet analysis to calculate power in 200 successive, overlapping time windows per trial. Analysis began with a 3-cycle wavelet and slowly expanded across the frequency spectrum in order to ensure that the shape of the analysis windows was controlled across the time and frequency spectrum. The number of cycles increased until the highest frequency where it reached 0.8 the number of cycles that would occur in a similar fast Fourier transformed window.

2.1.5 *Data analysis*

2.1.5.1 Whole-group analyses

2.1.5.1.1 Behavioral performance

Behavioral performance on the language analytic ability, working memory, grammar proficiency, and reading experience tasks was correlated to determine the degree to which predictor variables were related to one another. To assess behavioral performance on the sentence judgment task, a-prime scores were computed for all conditions using the *dprime* function in the *psycho* package in R (Makowski, 2018; R Core Team, 2019). A-prime is a non-parametric version of d-prime, a measure of sensitivity in Signal Detection Theory, that computes the differences between hits and false alarms in the data. An a-prime score of 0.5 indicates chance performance and a score of 1.0 indicates perfect performance. For syntax condition, hits included correct acceptance of grammatical conditions and correct rejections of ungrammatical conditions, and false alarms included incorrect rejections of grammatical conditions and incorrect acceptance of ungrammatical conditions. A one-way repeated-measures ANOVA was carried out to test whether sensitivity differed across syntactic constructions using condition as a within-subjects factor with 3 levels (word order, verb tense, collective noun). For the semantic block, a one-way repeated-measures ANOVA was first computed to test whether accuracy differed across the three conditions (expected, unexpected, implausible). Then, a-prime scores were computed separately for implausible and unexpected continuations relative to expected continuations. A pairwise t-test was computed to determine whether a-prime scores differed between implausible and unexpected conditions.

2.1.5.1.2 ERP analyses

ERPs were computed by averaging voltage in the N400 window from 300-500 ms post stimulus and the P600 window from 500-800 ms post stimulus across correct trials in all conditions, baseline corrected relative to a 100 ms prestimulus interval (-100-0 ms). For whole-group analyses, omnibus repeated-measures ANOVAs were carried out using the *anova_test* function from the *rstatix* R package (Alboukadel, 2020). For syntactic conditions and semantic conditions in midline and lateral electrodes separately. Midline electrodes contained Fz, Cz, and Pz. Lateral electrodes were grouped into four ROIs comprised of frontal and posterior electrodes in left and right hemispheres, each containing four electrodes: left frontal left frontal (F7, F3, FC5, FC1), right frontal (F8, F4, FC6, FC2), left posterior (CP5, CP1, P7, P3), and right posterior (CP6, CP2, P8, P4). For syntactic conditions, midline analyses contained 3 within-subject factors: grammaticality (grammatical, ungrammatical), condition (word order, verb tense, collective noun), and anteriority, each level corresponding to one channel (Fz: frontal, Cz: central, Pz: posterior). Lateral analyses contained two levels of anteriority (frontal, posterior) and an additional within-subjects factor of hemisphere (left, right). For semantic conditions, midline analyses contained 2 within-subjects factors: condition (expected, unexpected, implausible) and anteriority (Fz: frontal, Cz: central, Pz: posterior). Lateral analyses in the semantic conditions contained two levels of anteriority (frontal, posterior) and an additional within-subjects factor of hemisphere (left, right). Because individual differences in response to violations are of primary interest, discussion of group-level results will focus on the main effects of grammaticality and condition and any interactions containing these factors. Greenhouse-Geisser correction for inhomogeneity of variance was used for any repeated-measures factors with more than one degree of freedom in the numerator. In such cases, the corrected *p*-value is reported.

2.1.5.1.3 Time-frequency analyses

Previous research on beta frequency changes in response to syntactic and semantic violations differs considerably with regard to the analysis time window used. In order to characterize frequency changes in an unbiased manner, group-level analyses were conducted to determine windows of significance before testing for individual differences. Within each violation type, mean power in the -200 ms to 0 ms baseline was computed across conditions (e.g., grammatical vs. ungrammatical in syntax violations, expected vs. unexpected vs. implausible in the semantic violations). This mean baseline power was then subtracted from each condition such that power differences between conditions could not be due to differences in the baseline. Group-level significance was determined using cluster-corrected Monte Carlo permutation analyses provided by the Fieldtrip toolbox for EEGLab (Bastiaansen et al., 2010; Maris & Oostenveld, 2007; Oostenveld et al., 2011; Schneider & Maguire, 2018) in electrodes contained in the RDI within an 8 Hz to 40 Hz frequency range. This frequency range was chosen because it contains the traditionally defined beta frequency band (12 Hz to 30 Hz) with enough flexibility to capture individual differences and allow for cluster correction that is not affected by constricting the frequency range.

In these analyses, a test statistic was computed for every frequency and time point across conditions (*t* values for comparing the two levels in each syntax condition and *F* values for comparing the three levels in the semantic condition). Clusters were delineated by finding all data sample points that were significant at an alpha level of 0.05 and were clustered together according to proximity in time and frequency. Then, a cluster-level statistic was created for the observed data by summing the individual *t* or *F* statistics for that cluster. To create a null distribution against which to test the cluster-level statistic, subject averages in each electrode were randomly assigned to each of the experimental conditions and cluster-level statistics were computed from this randomized data. This step was repeated 10,000 times and a null distribution was created by extracting the largest cluster-level statistic from each permutation. Finally, the *p*-value of each observed cluster-level statistic was computed based on the permuted null distribution and observed clusters with *p*-values less than .05 were marked as significant.

Cluster-corrected permutation tests were conducted separately for each electrode in the ROI and the start and end points of the time window during which significance were extracted from each electrode that showed significant differences in the beta frequency range. The median start and end times across channels were used to determine the analysis window for extracting individual beta frequency power for each condition. Time windows of analysis for each condition were as follows: Word Order: 500-1000 ms; Verb Tense: 600-1200 ms; Collective Noun: 600-1200 ms; Semantic: 600-1200 ms. Average beta across ROI channels was then computed within the significant window in each condition for each participant.

For whole-group analyses, repeated-measures ANOVAs were carried out for syntax and semantic conditions separately. In syntax conditions, there were two within-subject factors with two levels each: grammaticality (grammatical, ungrammatical) and condition (word order, verb tense, collective noun). In semantic conditions, a one-way repeated measures ANOVA was computed with condition as a factor with three levels (expected, unexpected, implausible). Greenhouse-Geisser correction for inhomogeneity of variance was used for any repeated-measures factors with more than one degree of freedom in the numerator. In such cases, the corrected *p*-value is reported.

2.1.5.2 Individual differences analyses

For individual differences analyses, N400 and P600 effect magnitudes were computed within the centro-parietal ROI corresponding to the RDI. In syntactic conditions, N400 and P600 effects were computed by comparing grammatical and ungrammatical conditions (N400: grammatical - ungrammatical; P600: ungrammatical - grammatical). In semantic conditions, N400 and P600 magnitudes were computed in unexpected and implausible conditions relative to expected conditions (N400: unexpected/implausible - expected; P600: expected - unexpected/implausible). Correlation analyses were carried out to test for a negative relationship between N400 and P600 effect magnitudes previously observed in related literature (Grey et al., 2017; Tanner, 2019; Tanner et al., 2013, 2014; Tanner & Van Hell, 2014). A strong negative correlation indicates that there is good variability in individual ERP responses and that individual response patterns are prominently dominated by N400 or P600 effects. RDI values were computed to determine the degree to which a participant was N400- or P600-dominat in each condition using Equation 1:

$$\mathrm{RDI} = \frac{(P600\ Effect\ Magnitude) - (N400\ Effect\ Magnitude)}{\sqrt{2}} \qquad (1)$$

The RDI computes a participants' perpendicular least squares distance from an equal ERP effects magnitude line (see Figure 9 for an example).

A one-way repeated measures ANOVA was used to test whether variance in RDI values differed across syntactic and semantic conditions with the expectation that RDI values would pattern together for the syntax conditions and would have more positive RDI values (more P600-dominant responses) than semantic conditions, which would have more negative RDI values (more N400-dominant responses). RDI values in each condition were also correlated with their respective a-prime scores as well as with scores on the other behavioral measures to test whether variability in ERP responses could be explained by sensitivity to correct language use or factors such as working memory and language experience that have previously been implicated in individual differences in language processing.

Beta desynchronization was computed for each participant within the ROI for syntactic and semantic continuations separately. In syntax violations, magnitude of beta desynchronization was calculated by subtracting average beta in ungrammatical conditions from grammatical conditions for each syntactic condition. In semantic violations, magnitude of beta desynchronization was calculated by subtracting average beta in implausible or unexpected continuations from beta in expected continuations.

Prior to computing analyses testing the predictions of the Top-Down Prediction and Syntactic Sensitivity accounts, RDI values in the syntax conditions were related to each other in order to assess whether individual differences in grammar processing can be understood of as a single trait, as previously argued (Tanner, 2019; Tanner & Van Hell, 2014). First, RDI values in the syntax conditions were correlated with one another to determine if ERP response patterns were similarly N400- or P600-dominant across conditions. Additionally, a confirmatory factor analysis was carried out to test whether variability across syntactic conditions loaded onto one latent variable related specifically to variability in syntactic processing.

The three primary analyses proposed to address the predictions of the Top-Down Prediction and Syntactic Sensitivity accounts outlined in the introduction were carried out using regression analyses described below. All regression analyses were computed using the *lm* function in the *stats* package in R (R Core Team, 2019; Venables & Riple, 2002). When significant regression analyses did not meet assumptions of linear models, bisquare robust linear models were computed to test significance of predictors using the *rlm* function in the *MASS* package in R (Venables & Riple, 2002). In such cases, both linear and robust linear models are reported and discussed. In addition, all regression analyses were computed with and without language experience variables (English grammar proficiency and reading experience) and working memory scores. With the exception of one analysis, results were qualitatively similar with and without these variables. Given that fewer participants had grammar proficiency and reading experience scores, the simple models excluding these variables are reported to preserve power. For the analysis that differed, both models are reported and discussed.

2.1.5.2.1 Is variability in syntactic processing related to variability in language analytic ability?

To test the predictions of the Syntactic Sensitivity account that variability in grammatical processing is related to variability in language analytic ability, language analytic ability scores were regressed on RDI values in the word order, verb tense, and collective noun conditions.

2.1.5.2.2 Does variability in syntactic processing reflect variability in the nature of individual's linguistic predictions in language processing?

To test the predictions of the Top-Down Prediction account, two sets of analyses were conducted to address both features of prediction, *type* (RDI) and *strength* (beta). The goal of the first was to determine whether variability in syntactic processing and semantic processing were related with regard to the *type* of linguistic predictions made. Separate analyses were run with implausible and unexpected RDI values as outcome variables and word order, verb tense, and collective noun RDI values as predictors. The goal of the second set of analyses was to determine whether variability in syntactic processing and semantic processing were related with regard to the *strength* of linguistic predictions. Separate analyses were conducted using implausible and unexpected beta desynchronization as outcome variables and word order, verb tense, and collective noun beta desynchronization as predictors.

2.1.5.2.3 Does variability in type of linguistic predictions made during syntactic processing relate to variability in strength of linguistic predictions?

To further test the predictions of the Top-Down Prediction account, two sets of analyses were conducted to determine whether variability in the *type* of linguistic predictions individuals made in response to syntactic violations (RDI) were related to the *strength* of their linguistic predictions (beta). The goal of the first analysis was to determine whether variability in the type of linguistic predictions made in syntactic conditions was related to strength of linguistic predictions made in those conditions. For each syntax condition, RDI values were correlated with degree of beta desynchronization. The goal of the second analysis was to determine whether variability in the type of linguistic predictions made in syntactic conditions was related to strength of linguistic predictions made in semantic conditions. Separate analyses were run with implausible and unexpected beta desynchronization as outcome variables and word order, verb tense, and collective noun RDI values as predictors.

2.2 Experiment 2

2.2.1 *Participants*

Sixty-eight healthy adults aged 18-31 (M = 19.56, SD = 1.93; 45 female) were recruited for participation from the University of Washington Psychology Subject Pool. Socio-demographic variables relating to age, years of education, socio-economic status, language background and fluency in English and other languages, and handedness were collected for all participants. All participants were native Mandarin Chinese speakers who were born in a Mandarin Chinese-speaking country and began acquiring Mandarin Chinese at birth. Participants had an average of 10.69 years of English experience, defined as the non-overlapping sum of number of years of formal classroom instruction and number of years of residence in the United States (SD = 3.36). Five participants were removed from analyses because they reported having started formal English instruction before the age of 5 and/or reported English as their dominant language. All participants provided informed consent according to the standards outlined by the University of Washington Institutional Review Board and were compensated for their time with course credit. Data from 50 participants (Age: M = 19.68, SD = 2.12; 33 female; Years English Experience: M = 10.21, SD = 3.30) who completed both EEG data collection sessions with no more than 25% trial rejection rate per session (**see Analyses**) was used for subsequent analyses.

2.2.2 *Materials*

Participants were presented with two blocks of sentences: one block of Chinese sentences and one block of English sentences. The order of the blocks was counterbalanced across subjects. Both blocks contained violations of word order, a rule that is shared across the two languages. In noun phrases in both languages, possessive pronouns (e.g., my/his/their) must precede the noun they possess. If the determiner erroneously follows the noun, it is considered a violation of the syntactic structure. The word order condition will thus be referred to as the similar or familiar L1/L2 rule.

The English block also contained a violation of verb tense, that is unique to English. Verb tense agreement does not exist in Chinese, which is a morphologically simple language. Thus, the verb tense condition represents a more complex syntactic rule that native Chinese speakers have to learn when acquiring English. The verb tense condition will thus be referred to as the unique L2 rule.

2.2.2.1 Chinese block stimuli

The Chinese block contained syntactic violations of word order. The word order condition was composed of 60 Chinese sentences containing pronoun + noun verb phrases (e.g., *I had a fight with my friend this morning*). Violations were constructed by reversing the order of the pronoun and noun (e.g., *I had a fight with friend *my this morning*) (see Table 2). To ensure that the violation occurred only at the point when the noun was ungrammatically followed by a pronoun, sentences were written such that they remained grammatical until the point of the pronoun. Sentences were separated according to natural character boundaries that preserved the intended meaning of the sentence as it was presented incrementally to participants. On average, sentences ranged from 7 to 12 increments (M = 9.12, SD = 1.56).

Table 2. Example Sentences in Chinese Block. The critical word is underlined.

Condition	Example
Word Order	今天 上午 我的 朋友 / 朋友 *我的 和 我 打架了.
	Today morning my friend / friend *my and I fight.
	I had a fight with my friend/friend *my this morning.
Semantic	刚刚 地上 掉的 词典 / ?意思 像是 小雷的.
	Just on floor drop dictionary / ?meaning looks like Xiaolei’s.
	The dictionary/?meaning just dropped on the floor looks like Xiaolei’s.

An additional 60 filler sentences were constructed and further turned into semantic violations by replacing a plausible noun with an implausible noun (see Table 2). Sentences ranged from 6 to 13 words (M = 9.4, SD = 1.59).

All sentences were composed and reviewed by three native Mandarin Chinese speakers who were instructed to ensure that all sentences remained syntactically and semantically well-formed and plausible until the presence of the violation. All items were further rated by 83 native Mandarin Chinese speakers who were born in a Mandarin Chinese-speaking country, began acquiring Mandarin Chinese at birth, and for whom Mandarin Chinese is their dominant language (e.g., spending 50% or more of their life speaking, reading, and hearing Mandarin Chinese). Participants provided informed consent and were compensated for their time with course credit. Each item was rated by up to 50 participants on a scale of 1 to 5 based on how implausible (1) or plausible (5) they found the sentence to be in Mandarin Chinese. Results were in line with expectations: well-formed sentences were rated 4.69 in the word order condition (SD = 0.27) and 4.64 in the semantic filler condition (SD = 0.23), while violations were rated 1.89 in the word order condition (SD = 0.36) and 1.52 in the semantic filler condition (SD = 0.23).

Two experimental lists of 120 sentence were constructed for the Chinese block containing 60 well-formed sentences (30 word order, 30 fillers), 30 word order violations, and 30 semantic violations. Each pseudorandomized list contained only one version of each experimental sentence such that no participant saw any given item in both experimental conditions. Participants were pseudorandomly assigned to one of the two stimulus lists.

2.2.2.2 English block stimuli

The English block contained two syntactic conditions, a word order condition and a verb tense condition. Stimuli were the same as those described in Experiment 1, with the exception that the word order condition was composed only of sentences containing pronoun + noun verb phrases (e.g., *Their niece ate her <u>soup</u> without spilling a drop*) to maximize similarity between languages. Additional sentences were constructed in each condition such that each condition had a total of 60 sentences. Violations were constructed from these sentences in the same manner described in Experiment 1. The word order condition ranged from 7 to 12 words (M = 9.68, SD

= 1.18) and the verb tense condition ranged from 7 to 11 words (M = 9.2, SD = 0.96).

Two experimental lists of 150 sentence were constructed for the English block containing 60 sentences from the word order and verb-agreement conditions each. Half of the sentences in each condition were grammatical and the other half were ungrammatical such that there were 30 items per each syntax by grammaticality condition. An additional 25 filler sentences contained semantic violations to vary the types of violations in the block (e.g., *Flowers need sunlight and ?<u>toys</u> to fully grow*). Each pseudorandomized list contained only one version of each experimental sentence such that no participant saw any given item in more than one experimental condition. Participants were pseudorandomly assigned to one of the two stimulus lists.

2.2.2.3 Behavioral tasks

Chinese participants completed all of the behavioral tasks described in Experiment 1 with the exception of the ART due to time constraints.

2.2.3 *Procedures*

Experimental procedures were the same as described in Experiment 1. Participants were presented with the Chinese block in one session and the English block in the other session. The other of the sessions was counterbalanced.

2.2.4 *EEG recording and processing*

All EEG analysis methods were the same as described in Experiment 1. For ERP averages, all participants not excluded for artifact rejection rates in excess of 25% had an average rejection rate of 5% of trials in the Chinese block (SD = 0.04) and 5% of trials in the English block (SD = 0.05).

2.2.5 *Data analysis*

2.2.5.1 Whole-group analyses

2.2.5.1.1 Behavioral performance

Behavioral performance on the language analytic ability, working memory, and English grammar proficiency tasks and year of English experience were correlated to determine the degree to which predictor variables were related to one another. A-prime scores were computed as described in Experiment 1 for Chinese word order, English word order, and English verb tense separately by comparing hits and false alarms in grammatical and ungrammatical conditions. A one-way repeated-measures ANOVA was carried out to test whether sensitivity differed across conditions.

2.2.5.1.2 ERP analyses

ERPs were computed as described in Experiment 1. For Chinese stimuli omnibus ANOVAs, midline analyses contained 2 within-subject factors: grammaticality (grammatical, ungrammatical) and anteriority (Fz: frontal, Cz: central, Pz: posterior). Lateral analyses contained two levels of anteriority (frontal, posterior) and an additional within-subjects factor of hemisphere (left, right). For English stimuli omnibus ANOVAs, analyses contained 3 within-subject factors: grammaticality (grammatical, ungrammatical), condition (word order, verb tense), and anteriority, each level corresponding to one channel (Fz: frontal, Cz: central, Pz: posterior). Lateral analyses contained two levels of anteriority (frontal, posterior) and an additional within-subjects factor of hemisphere (left, right). Because individual differences in response to violations are of primary interest, discussion of group-level results will focus on the main effect of grammaticality and any interactions containing this factor.

2.2.5.1.3 Time-frequency analyses

Time-frequency analyses were computed as described in Experiment 1. Time windows for group-level and individual time-frequency analyses were as follows: Chinese word order: 600-1000 ms; English word order: 500-1000 ms; English verb tense: 700-1300 ms. For whole-group analyses, repeated-measures ANOVAs were carried out for Chinese and English conditions separately. For Chinese stimuli, there was one within-subjects factor of grammaticality (grammatical, ungrammatical). For English stimuli, there were two within-subject factors with two levels each: grammaticality (grammatical, ungrammatical) and condition (word order, verb tense).

2.2.5.2 Individual differences analyses

For individual differences analyses, N400 and P600 effect magnitudes were computed as described for in Experiment 1 for syntactic condition. Correlation analyses were carried out to test for a negative relationship between N400 and P600 effect magnitudes in the three conditions and RDI values were computed for each individual. Beta desynchronization was computed as described in Experiment 1 for syntactic conditions.

A one-way repeated measures ANOVA was used to test whether variance in RDI values differed across Chinese and English conditions. To determine if variability in ERP responses could be explained by sensitivity to correct language use or other previously implicated factors, RDI values in each condition were correlated with their respective a-prime scores as well as with working memory and English grammar proficiency scores and years of English experience.

The three primary analyses proposed to address the predictions of the Top-Down Prediction and Syntactic Sensitivity accounts outlined in the introduction were carried out using regression analyses described below. In addition, all regression analyses were computed with and without working memory scores. Results were qualitatively similar with and without working memory as a predictor and thus the simpler models excluding this variable are reported.

2.2.5.2.1 Is variability in L2 unique rule processing related to variability in language analytic ability?

To test the predictions of the Syntactic Sensitivity account that variability in grammatical processing is related to variability in language analytic ability, separate analyses were run with English verb tense RDI and a-prime scores as outcome variables and language analytic ability as a predictor variable, controlling for L2 language experience (English grammar proficiency, years of English experience).

2.2.5.2.2 Does variability in L2 unique rule processing reflect variability in the nature of individual's linguistic predictions in language processing?

To test the predictions of the Top-Down Prediction account, two sets of analyses were conducted to address both features of prediction, *type* (RDI) and *strength* (beta). The goal of the first set of analyses was to determine whether variability in L2 unique rule processing and L1 processing were related with regard to the *type* of linguistic predictions made, over and above L2 language experience. Separate analyses were carried out with English verb tense RDI and a-prime scores as outcome variables and Chinese word order RDI as a predictor variable, controlling for L2 language experience (English word order RDI, English grammar proficiency, years of English experience). The goal of the second set of analyses was to determine whether variability in L2 unique rule processing and L1 processing were related with regard to the *strength* of linguistic predictions. Separate analyses were conducted with English verb tense beta desynchronization and a-prime scores as outcome variables and Chinese word order beta desynchronization as a predictor variable, controlling for L2 language experience (English word order beta desynchronization, English grammar proficiency, years of English experience).

2.2.5.2.3 Does variability in type of linguistic predictions made during syntactic processing relate to variability in strength of linguistic predictions?

To further test the predictions of the Top-Down Prediction account, two analyses were conducted to determine whether variability in the *type* of linguistic predictions individuals made in response to the L2 unique rule (RDI) was related to the *strength* of their linguistic predictions (beta). The goal of the first analysis was to determine whether variability in the type of linguistic predictions made in response to the L2 unique rule was related to strength of linguistic predictions made in that condition, controlling for L2 language experience. English verb tense RDI values were regressed on English verb tense beta desynchronization, English grammar proficiency, and years of English experience. The goal of the second analysis was to determine whether variability in the type of linguistic predictions made in response to the L2 unique rule was related to strength of linguistic predictions made in L1, controlling for L2 language experience. English verb tense RDI values were regressed on Chinese word order beta desynchronization, English verb tense beta desynchronization, English grammar proficiency, and years of English experience.

Chapter 3. RESULTS

3.1 Experiment 1

3.1.1 *Whole-group results*

3.1.1.1 Behavioral results

There was a good deal of variability in performance on behavioral tasks across participants. Table 3 contains descriptive statistics for language analytic ability, working memory, grammar proficiency, and reading experience and Figure 1 provides distributions of scores. There was only one significant correlation between language analytic ability and reading experience, *r*(39) = 0.33, uncorrected *p* = 0.035.

Table 3. Means, standard deviations, and ranges of scores on behavioral tasks in Experiment 1.

	Mean	**SD**	**Range**
Language Analytic Ability	67.19	21.44	0-100
Working Memory	23.58	12.75	2-61
Grammar Proficiency	0.91	0.04	0.8-0.99
Reading Experience	6.15	4.85	1-19

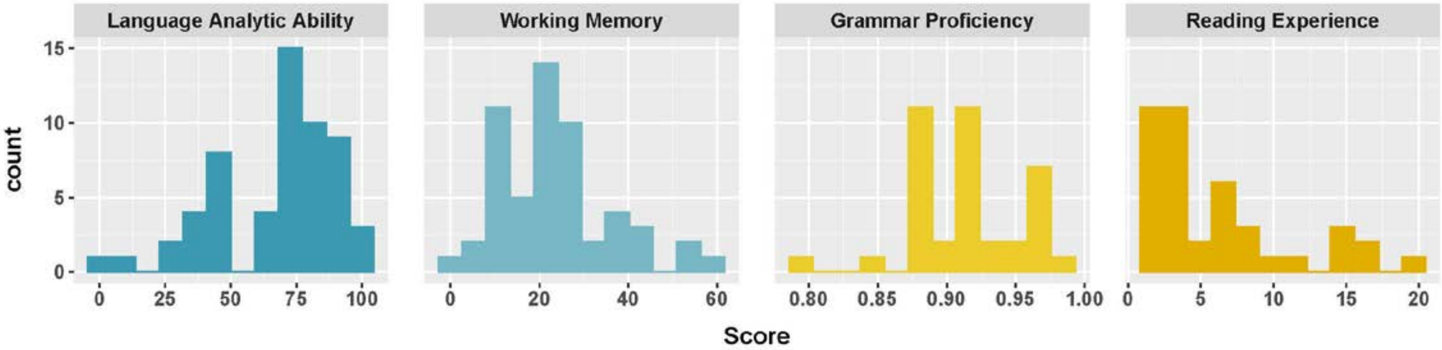

Figure 1. Distributions of scores for behavioral tasks in Experiment 1.

Participants generally had high behavioral performance on the sentence judgment tasks (see Figure 2). There was a significant difference in a-prime scores across syntactic conditions, *F*(1.25, 69.97) = 41.27, *p* < .001. FDR-corrected pairwise t-tests showed that a-prime in the word order (M = 0.98, SD = 0.02) and verb tense conditions (M = 0.98, SD = 0.02) was significantly larger than a-prime in collective noun condition (M = 0.93, SD = 0.05).

A paired samples t-test showed that the unexpected continuation (M = 0.85, SD = 0.1) had a significantly lower a-prime score than the implausible continuation (M = 0.97, SD = 0.02), *t*(56) = 9.53, *p* < .001. Comparing behavioral performance across all three semantic conditions, there was also a significant difference in accuracy across conditions, *F*(1.29, 72.21) = 160.61, *p* < .001. FDR-corrected pairwise t-tests showed that accuracy was highest in the implausible condition (M = 0.97, SD = 0.06), followed by the expected condition (M = 0.92, SD = 0.07), and finally the unexpected condition (M = 0.62, SD = 0.18).

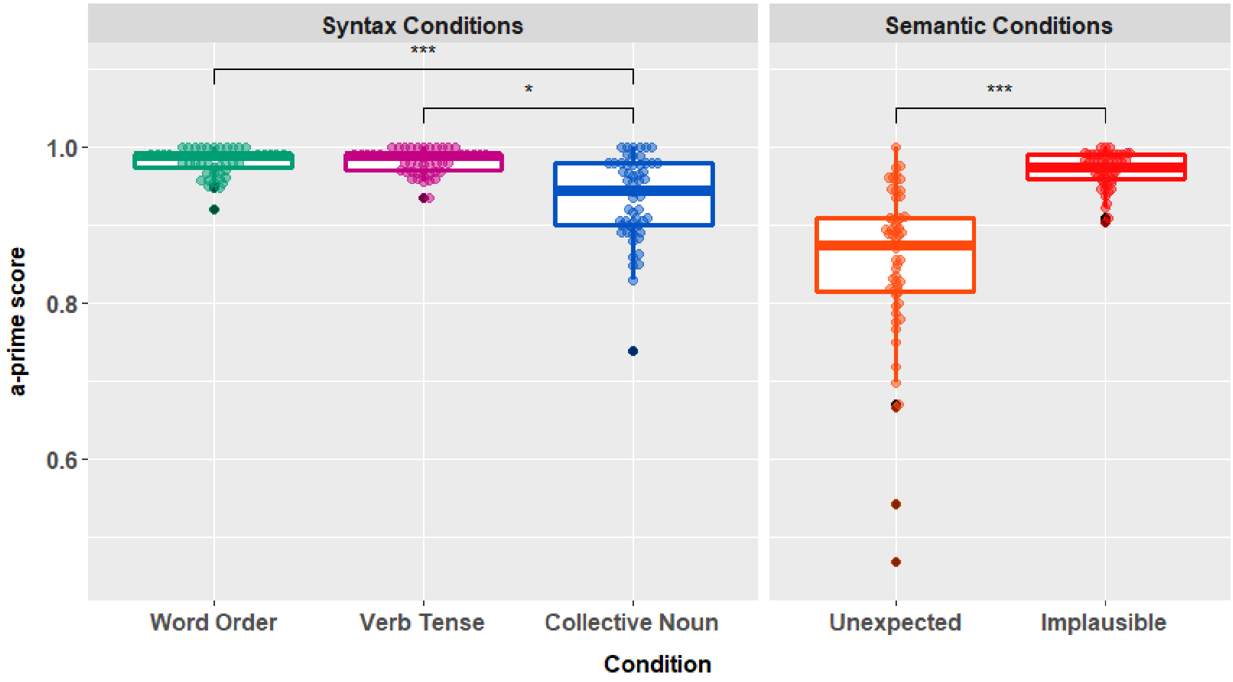

Figure 2. Boxplots of a-prime scores by condition in Experiment 1.

3.1.1.2 ERP results

3.1.1.2.1 Syntax conditions

Grand mean waveforms for the three syntax conditions are shown in Figure 3, Figure 4, and Figure 5. Visual inspection of the waveforms showed that there was a widespread P600 effect elicited maximally in the centro-parietal regions in all three syntax conditions. The verb tense condition also showed an N400 effect that peaked in the fronto-central electrodes.

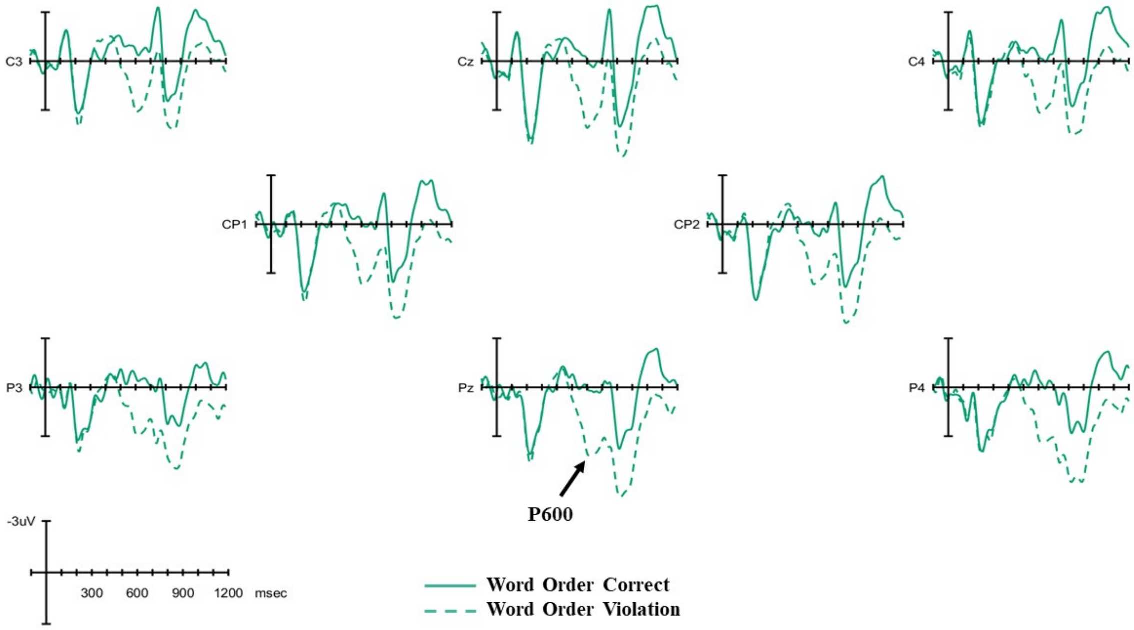

Figure 3. Grand mean ERP waveforms for the word order condition in Experiment 1. Eight representative electrodes corresponding to the ROI for the RDI are plotted. The mean activity for the grammatical condition is indicated by the solid line and the dashed line indicates mean activity for the ungrammatical condition. The onset of the critical word is indicated by the -3μV vertical calibration bar. Each tick mark represents 100 ms of time. Negative voltage is plotted up.

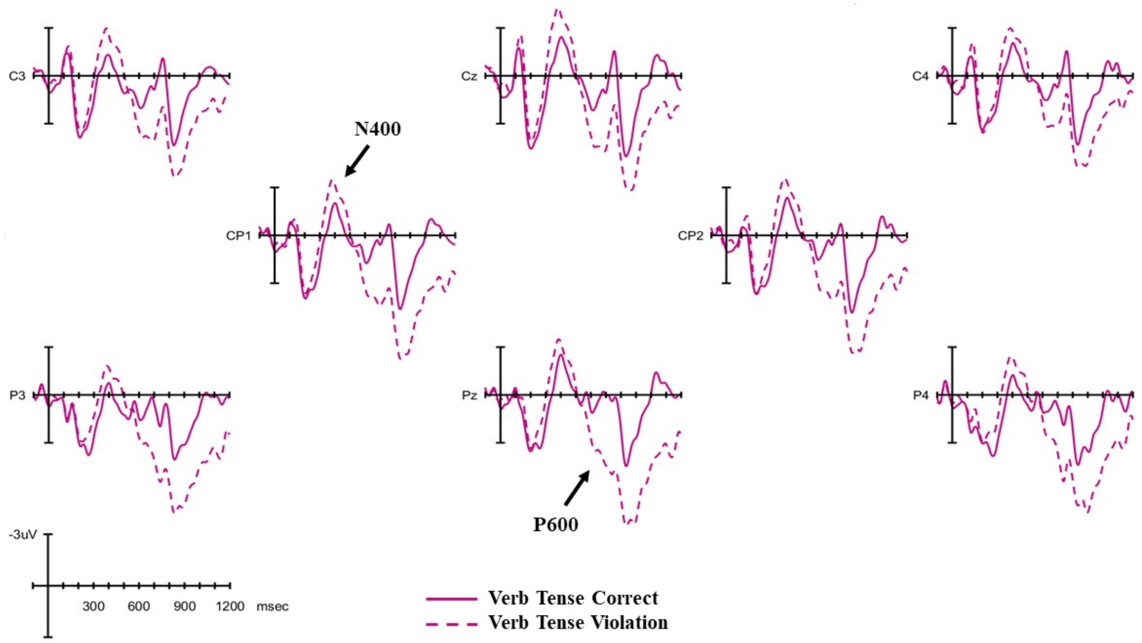

Figure 4. Grand mean ERP waveforms for the verb tense condition in Experiment 1. The mean activity for the grammatical condition is indicated by the solid line and the dashed line indicates mean activity for the ungrammatical condition.

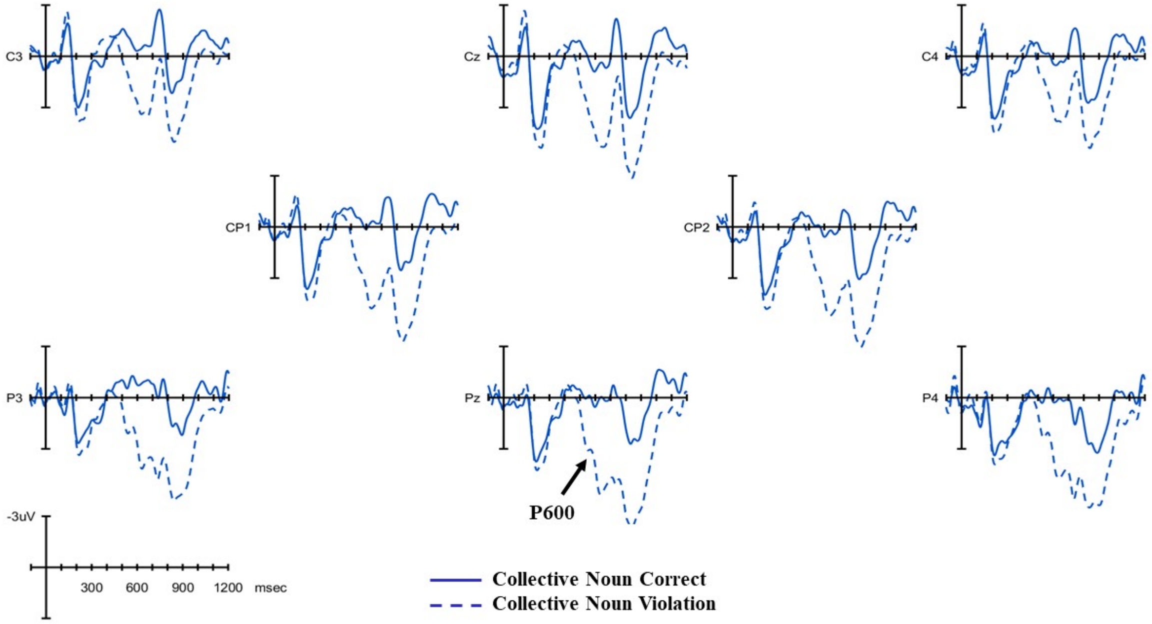

Figure 5. Grand mean ERP waveforms for the collective noun condition in Experiment 1.

The mean activity for the grammatical condition is indicated by the solid line and the dashed line indicates mean activity for the ungrammatical condition.

Table 4 reports the results of the omnibus ANOVA for the three syntax conditions. A grammaticality by condition interaction in the N400 window in midline and lateral electrode sites confirmed the presence of an N400 effect in the verb tense condition. FDR-corrected pairwise t-tests showed that the difference in voltage between grammatical and ungrammatical conditions were larger for the verb tense condition (Midline: M = 1.77 µV, SE = 0.31; Lateral: M = 1.52 µV, SE = 0.22) than in the word order (M = -0.16 µV, SE = 0.21; Lateral: M = -0.07 µV, SE = 0.14) or collective noun conditions (M = -0.14 µV, SE = 0.26; Lateral: M = 0.1 µV, SE = 0.19). The verb tense condition drove a main effect of grammaticality in lateral sites and a grammaticality by hemisphere interaction showed that N400 magnitude was larger over left hemisphere electrodes (M = 0.67 µV, SE = 0.15) compared to right hemisphere electrodes (M = 0.36 µV, SE = 0.16). This interaction was further qualified by an interaction with anteriority. Follow-up ANOVAs showed that the grammaticality by hemisphere interaction was significant in frontal electrodes (*F*(1, 170) = 11.52, *p* < .001) but not posterior electrodes (*F*(1, 170) = 0.92, *p* = 0.338). These data show that the verb tense condition showed an N400 effect that was maximal over frontal electrodes in the left hemisphere.

Table 4. F-statistics from ANOVA results for midline and lateral electrodes in N400 and P600 windows for syntax conditions in Experiment 1. Degrees of freedom are noted next to the main effects and interactions. Significant results are bolded.

	N400 (300-500 ms)	**P600** (500-800 ms)
Midline		
Gram. (1, 56)	3.34	**74.64*****
Cond. (2, 112)	**16.87*****	**3.65***

Ant. (2, 112)	**4.86***	**30.05*****
Gram. x Cond. (2, 112)	**7.75*****	2.06
Gram. x Ant. (2, 112)	1.29	**18.64*****
Cond. x Ant. (4, 224)	**6.13****	**8.3*****
Gram. x Cond. x Ant. (4, 224)	1.50	0.6
Lateral		
Gram. (1, 56)	**6.56***	**74.85*****
Cond. (2, 112)	**3.96***	**9.69*****
Hem. (1, 56)	**10.87****	**26.85*****
Ant. (1, 56)	**36.91*****	**23.32*****
Gram. x Cond. (2, 112)	**6.97****	2.6
Gram. x Hem. (1, 56)	**7.34****	0.65
Cond. x Hem. (2, 112)	**18.36*****	**26.92*****
Gram. x Ant. (1, 56)	0.00	**6.20***
Cond. x Ant. (2, 112)	1.07	0.75
Hem. x Ant. (1, 56)	0.19	**5.15***
Gram. x Cond. x Hem. (2, 112)	1.49	0.03
Gram. x Cond. x Ant. (2, 112)	0.66	2.99
Gram. x Hem. x Ant. (1, 56)	**6.96***	3.26
Cond. x Hem. x Ant. (2, 112)	0.42	0.46
Gram. x Cond. x Hem. x Ant. (2, 112)	1.32	0.97

Note. Gram. = Grammaticality; Cond. = Condition; Hem. = Hemisphere; Ant. = Anteriority. *$p < 0.05$; **$p < .01$; ***$p < 0.001$.

In the P600 window, there was a widely distributed main effect of grammaticality across midline and lateral electrodes, such that ungrammatical conditions had larger magnitudes (Midline: M = 2.85 µV, SE = 0.17; Lateral: M = 1.95 µV, SE = 0.11) relative to grammatical conditions (Midline: M = 0.02 µV, SE = 0.13; Lateral: M = -0.03 µV, SE = 0.09). The magnitude of the P600 effect did not differ across syntactic conditions. There was an additional grammaticality by anteriority interaction that indicated the difference between grammatical and ungrammatical conditions was larger in posterior sites (Midline: M = 3.56 µV, SE = 0.33; Lateral: M = 2.28 µV, SE = 0.19) compared to frontal sites (Midline: M = 1.86 µV, SE = 0.31;

Lateral: M = 1.69 µV, SE = 0.18).

Overall, results support the observed effects in the grand mean ERP waveforms that all syntax conditions showed P600 effects that were maximal in posterior electrodes, which is typical for this ERP component. The verb tense condition additionally showed an N400 effect that was maximal over left frontal electrodes, which is characteristic of the LAN component. However, the LAN has been argued to be an artifact that results from averaging N400 and P600 magnitudes together (Tanner, 2015). The large magnitude of the P600 masks the N400 in central electrodes where the two components tend to be maximal and thus the N400 appears to peak in left frontal electrodes where the P600 component does not have as strong of an influence in the average.

3.1.1.2.2 Semantic conditions

Grand mean waveforms for the three semantic conditions are shown in Figure 6. Visual inspection of the waveforms showed that there was a widespread N400 effect in both unexpected and expected conditions that was elicited maximally in centro-parietal regions. In the unexpected condition, there was an additional positive-going component in the 500-800ms window that was maximal in frontal electrodes resembling the late frontal positivity. In the implausible condition, there was an additional P600 effect evident in centro-parietal electrodes.

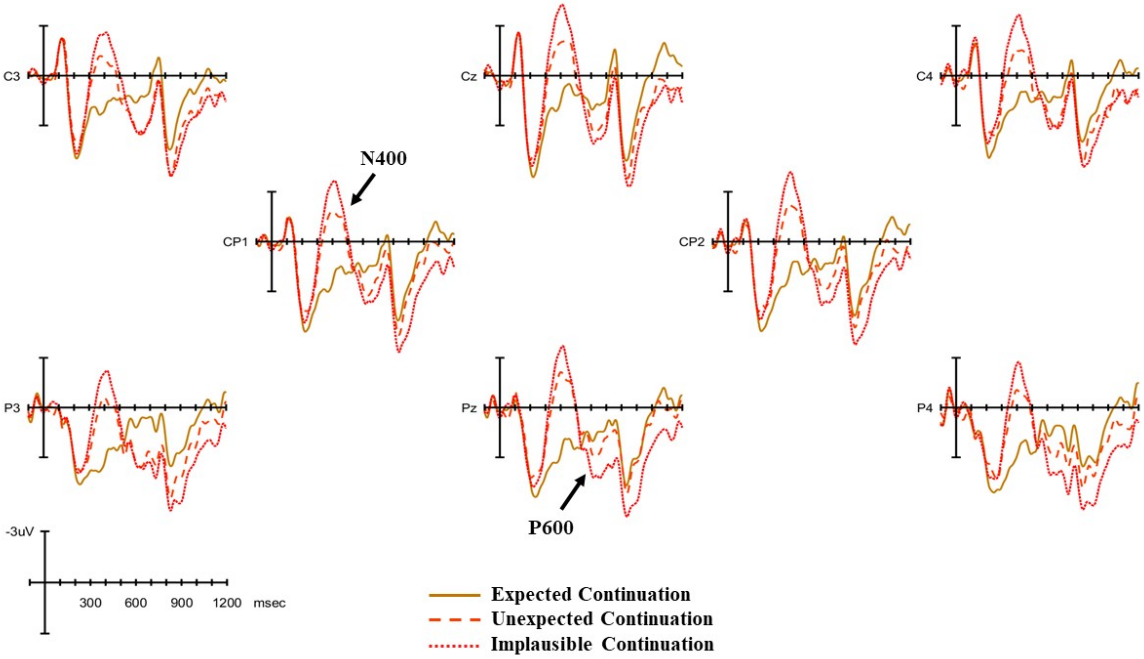

Figure 6. Grand mean ERP waveforms for the semantic conditions in Experiment 1. The mean activity for the expected condition is indicated by the solid line, mean activity for the unexpected condition is indicated by the dashed line, and mean activity for the implausible condition is indicated by the dotted line.

Table 5 reports the results of the omnibus ANOVA for the three semantic conditions. A main effect of condition in the N400 window across midline and lateral electrode sites confirmed the presence of the N400 effect. FDR-corrected pairwise t-tests showed that implausible N400 magnitude (Midline: M = -2.07 µV, SE = 2.84; Lateral: M = -1.14 µV, SE = 2.33) was smaller than unexpected N400 magnitude (Midline: M = -0.72 µV, SE = 3.44; Lateral: M = 0.02 µV, SE = 2.9) and both were smaller than expected N400 magnitude (Midline: M = 1.6 µV, SE = 3.51; Lateral: M = 1.63 µV, SE = 2.81). There was an additional condition by anteriority interaction in midline and lateral electrodes that indicated N400 effect was larger in posterior electrodes in unexpected (Midline: M = 3.69 µV, SE = 0.56; Lateral: M = 2.42 µV, SE = 0.36) and implausible conditions (Midline: M = 4.62 µV, SE = 0.48; Lateral: M = 3.42 µV, SE = 0.27) compared to frontal electrodes in the unexpected (Midline: M = 0.47 µV, SE = 0.6; Lateral: M = 0.81 µV, SE = 0.32) and implausible conditions (Midline: M = 2.13 µV, SE = 0.45; Lateral: M = 2.14 µV, SE = 0.27). A condition by hemisphere interaction in lateral electrodes also showed that the N400 effect was larger in both conditions in right hemisphere (Unexpected: M = 1.89 µV, SE = 0.37; Implausible: M = 3.02 µV, SE = 0.28) than in the left hemisphere (Unexpected: M = 1.34 µV, SE = 0.34; Implausible: M = 2.53 µV, SE = 0.27).

Table 5. F-statistics from ANOVA results for midline and lateral electrodes in N400 and P600 windows for semantic conditions in Experiment 1. Degrees of freedom are noted next to the main effects and interactions. Significant results are bolded.

	N400 (300-500 ms)	**P600** (500-800 ms)
Midline		
Cond. (2, 112)	**32.42*** **	**5.78** **
Ant. (2, 112)	**5.99** **	**7.58** **
Cond. x Ant. (4, 224)	**23.44*** **	**12.69*** **
Lateral		
Cond. (2, 112)	**25.99*** **	**6.17** **
Hem. (1, 56)	3.90	0.65
Ant. (1, 56)	**16.03*** **	**29.99*** **
Cond. x Hem. (2, 112)	4.79	**6.45** **
Cond. x Ant. (2, 112)	**19.88*** **	**14.83*** **
Hem. x Ant. (1, 56)	**4.14* **	**14.44*** **
Cond. x Hem. x Ant. (2, 112)	0.08	2.39

Note. Cond. = Condition; Hem. = Hemisphere; Ant. = Anteriority.

**p* < 0.05; ***p* < .01 , ****p* < 0.001.

A main effect of condition in the 500-800ms window across midline and lateral electrode sites confirmed the presence of a late positivity, which could be interpreted as a P600 effect or a late frontal positivity. FDR- corrected pairwise t-tests showed that implausible (Midline: M = 2.45 µV, SE = 3.31; Lateral: M = 1.89 µV, SE = 2.82) and unexpected magnitudes (Midline: M = 2.21 µV, SE = 3.47; Lateral: M = 2.1 µV, SE = 2.98) were both larger (more positive) than in the expected condition (Midline: M = 0.88 µV, SE = 2.77; Lateral: M = 0.71 µV, SE = 2.24), although they did not differ from each other. A condition by anteriority interaction in midline and lateral electrodes showed opposite patterns for unexpected and implausible conditions with regard to anteriority that clarified which late positivity effect occurred in each condition. In the unexpected condition, the late positivity was larger in the frontal electrodes in the midline (M = 1.59 µV, SE = 0.34) compared to posterior electrodes (M = 1.18 µV, SE = 0.36), but this difference was not significant in the lateral electrodes. This pattern is consistent with the interpretation of the positive-going component in the 500-800ms as the late frontal positivity that has been observed to follow the N400 effect in unexpected but plausible semantic continuations. In the implausible condition, on the other hand, the late positivity followed the distribution of the P600 effect, which has been observed to follow the N400 effect in implausible semantic continuations. P600 magnitude did not differ between frontal and posterior electrodes in the midline, but the P600 was larger in lateral posterior sites (M = 1.76 µV, SE = 0.3) compared to lateral frontal site (M = 0.6 µV, SE = 0.27). Finally, there was a condition by hemisphere interaction in the 500-800ms window as well that showed the late positivity was larger in the left hemisphere (M = 1.79 µV, SE = 0.32) compared to the right hemisphere (M = 0.99 µV, SE = 0.37). The magnitude of the P600 effect in the implausible condition did not differ across hemisphere.

Overall, results support the observed effects in the grand mean ERP waveforms that both implausible and unexpected conditions elicited more negative N400 magnitudes than the expected condition as well as late positivities. The posteriorly maximal distribution of the late positivity in the implausible condition corresponded to the typical scalp distribution of the P600 effect. The late positivity in the unexpected condition corresponded to the late frontal positivity in frontal electrodes. However, since the magnitude of the late positivity did not differ with anteriority in the lateral electrodes, it is possible that the late positivity in the unexpected condition could contain signatures of the P600 as well as the late frontal positivity. This ensures that the variability in unexpected continuation processing captured in the RDI does indeed reflect differences in P600 magnitude and not the late frontal positivity as it is calculated using centro-posterior electrodes.

3.1.1.3 Time-frequency results

3.1.1.3.1 Syntax conditions

Mean beta power by grammaticality and condition in the syntax conditions are plotted in Figure 7. A two-way repeated-measures ANOVA analyzing differences in beta desynchronization across grammaticality and syntax conditions revealed a main effect of grammaticality ($F(1, 56) = 49.27$, $p < .001$) and a main effect of syntax condition ($F(2, 112) = 9.42$, $p < .001$), but no interaction between grammaticality and syntax condition ($p = 0.765$). Beta power was smaller in ungrammatical conditions (M = -0.56 dB, SE = 0.04) relative to grammatical conditions (M = -0.19 dB, SE = 0.04). FDR-corrected pairwise t-tests showed that the verb tense condition had more beta desynchronization across both grammaticality conditions (M = -0.57 dB, SE = 0.06) compared to the word order (M = -0.28 dB, SE = 0.05) and collective noun conditions (M = -0.27 dB, SE = 0.05), which did not differ from one another. Overall, ungrammatical conditions resulted in greater beta desynchronization relative to grammatical conditions, and the degree of beta desynchronization did not differ across syntactic conditions. This is consistent with the established literature that syntactic violations result in smaller beta values compared to syntactically well-formed stimuli. Thus, in all three syntactic conditions, violations interrupted the maintenance of the top-down cognitive set of the sentence meaning, and the degree of this interruption was similar across syntactic conditions.

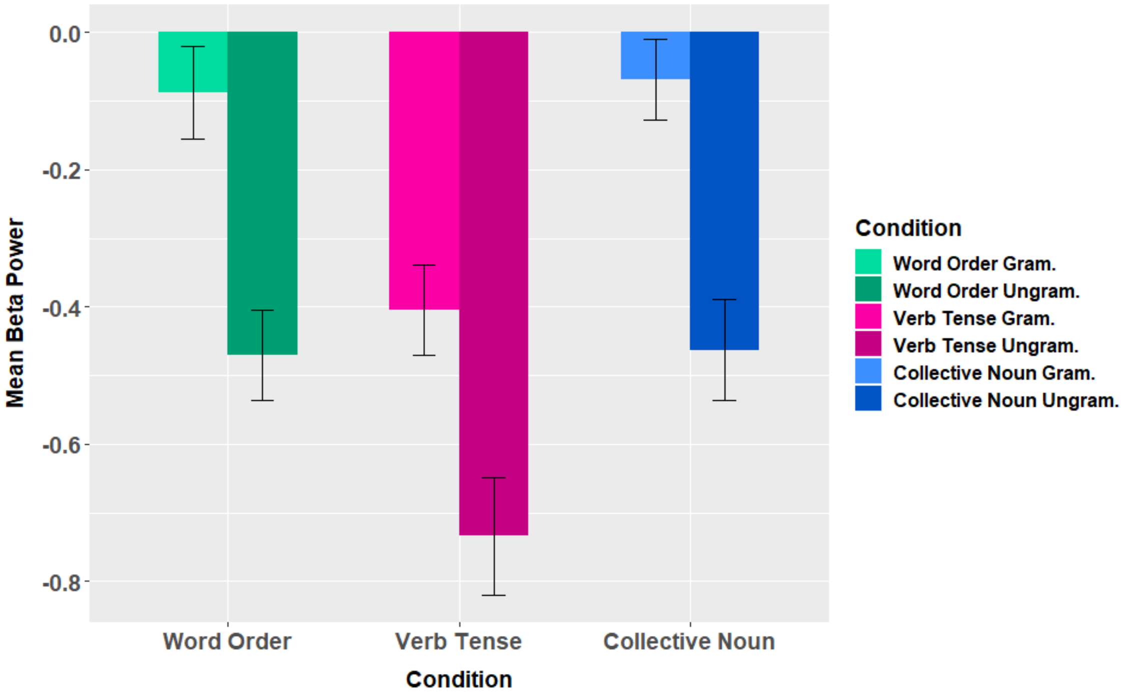

Figure 7. Mean beta power by grammaticality and condition in syntactic conditions in Experiment 1. Error bars indicate one standard error above and below the mean.

3.1.1.3.2 Semantic conditions

A one-way repeated-measures ANOVA analyzing differences in beta desynchronization across semantic conditions revealed that there was a significant difference in degree of beta desynchronization across the three semantic conditions, *F*(2, 112) = 21.53, *p* < .001). FDR-corrected pairwise t-tests showed that the implausible condition had more beta desynchronization (M = -0.67 dB, SE = 0.06) than either the expected (M = -0.32 dB, SE = 0.06) or unexpected conditions (M = -0.35 dB, SE = 0.06), which did not differ from one another. Thus, violations in the implausible condition interrupted maintenance of the top-down cognitive set of the sentence context, whereas violations in the unexpected condition did not.

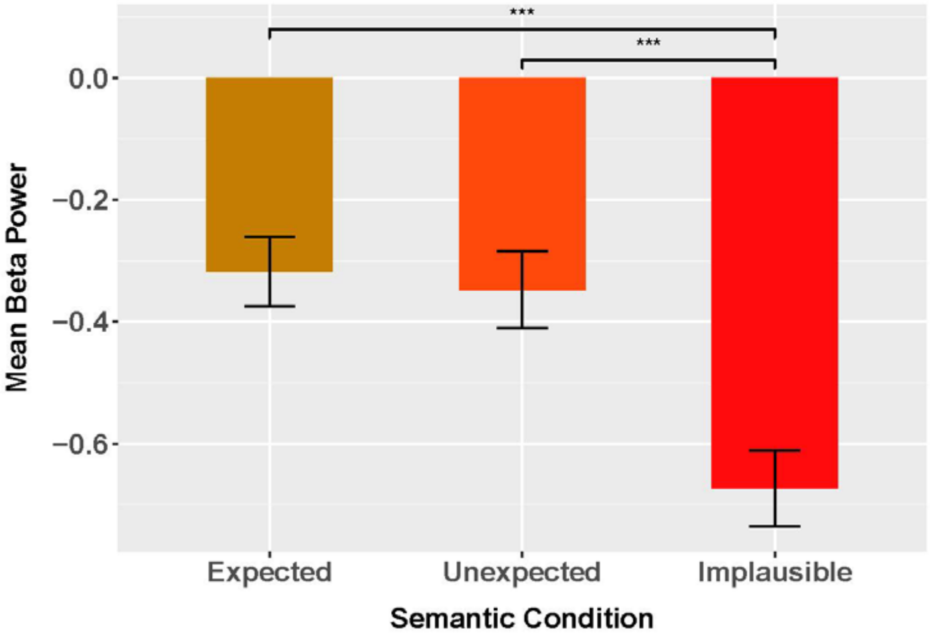

Figure 8. Mean beta power by semantic condition in Experiment 1. Error bars indicate one standard error above and below the mean. ***p < .001.

3.1.2 *Individual differences results*

In all syntax and semantic violation conditions, there was a good deal of variability in participant ERP response patterns. Figure 9 shows each participant's N400 and P600 magnitudes plotted against each other for each condition. There was a significant negative relationship between N400 effect magnitude and P600 effect magnitude in all conditions: word order: *r*(55) = -0.5, *p* < .001; verb tense: *r*(55) = -0.56, *p* < .001; collective noun: *r*(55) = -0.61, *p* < .001; unexpected: *r*(55) = -0.67, *p* < .001; implausible: *r*(55) = -0.28, *p* = 0.037. These results are in line with previous research indicating a tradeoff between N400 and P600 effects across a variety of linguistic anomalies (Grey et al., 2017; Kim et al., 2018; Meulman et al., 2014; Qi et al., 2017; Tanner, 2019; Tanner et al., 2014; Tanner & Van Hell, 2014). This further suggests that RDI values computed from these ERP effects capture the degree of this tradeoff across participants.

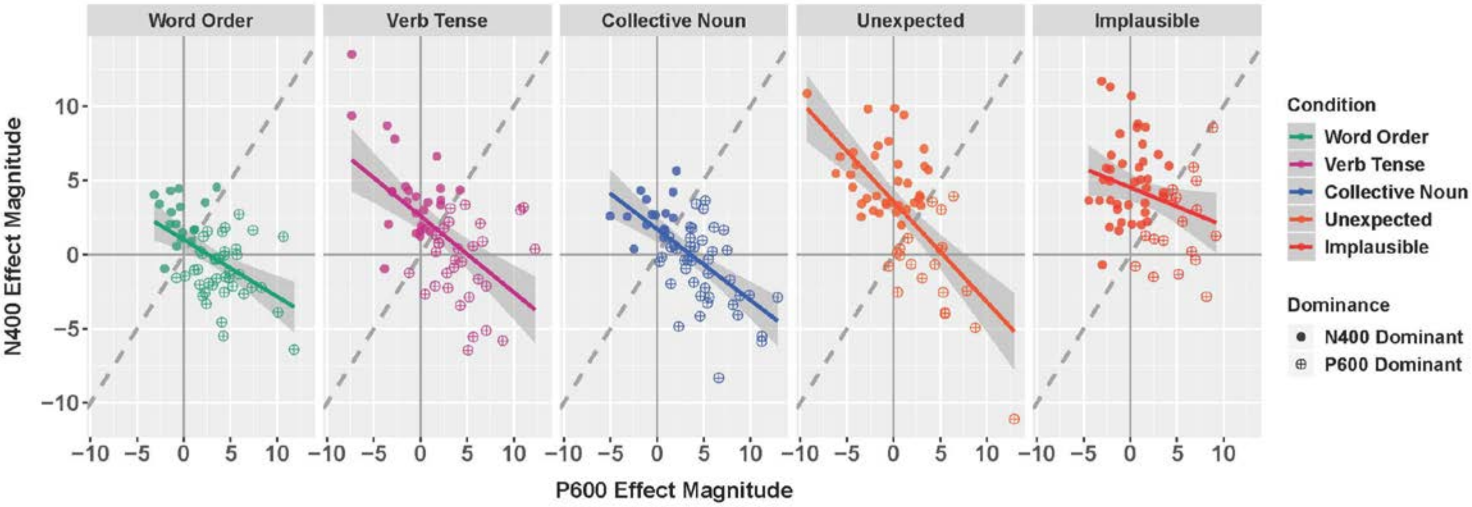

Figure 9. Participant N400 and P600 effect magnitudes by condition in Experiment 1. Dotted grey lines indicate line of equal ERP effect size. Solid circles to the left of the equal dominance line indicate N400-dominant individuals and crossed open circles the right of the equal dominance line indicate P600-dominant individuals. Solid color lines indicate regression lines.

Participants appeared to have more variable responses in the verb tense condition than in the word order and collective noun conditions. The semantic violation conditions also appeared to have more N400- dominant individuals than in the syntax conditions. Table 6 shows a summary of mean RDI values and proportion of P600-dominant individuals by condition and Figure 10 shows boxplots with overlaid participant RDI values by condition. The proportion of P600-dominant individuals in the word order and collective noun conditions was fairly high (M = 0.72) compared to the semantic conditions where the proportion was fairly low (M = 0.34). The proportion of P600-dominant individuals in the verb tense condition was in the middle (M = 0.54). A repeated-measures ANOVA revealed that RDI values across conditions were significantly different from each other, *F*(4, 224) = 14.22, *p* < .001. FDR- corrected pairwise t-tests showed that only word order and collective noun RDI values and unexpected and implausible RDI values were not significantly different from one another. These results confirmed that participants were predominantly P600-dominant in both word order and collective noun conditions and N400-dominant in semantic conditions, but were more variable in the verb tense condition, falling somewhere in between.

Table 6. Means and standard deviations of RDI values and proportion of P600-dominant individuals by condition.

	RDI Mean (SD)	**Proportion P600-Dominant**
Syntactic Conditions		
Word Order	2.15 (3.54)	0.70
Verb Tense	0.42 (4.88)	0.54
Collective Noun	2.59 (4.15)	0.74
Semantic Conditions		
Unexpected	-1.44 (5.18)	0.35
Implausible	-1.80 (3.73)	0.32

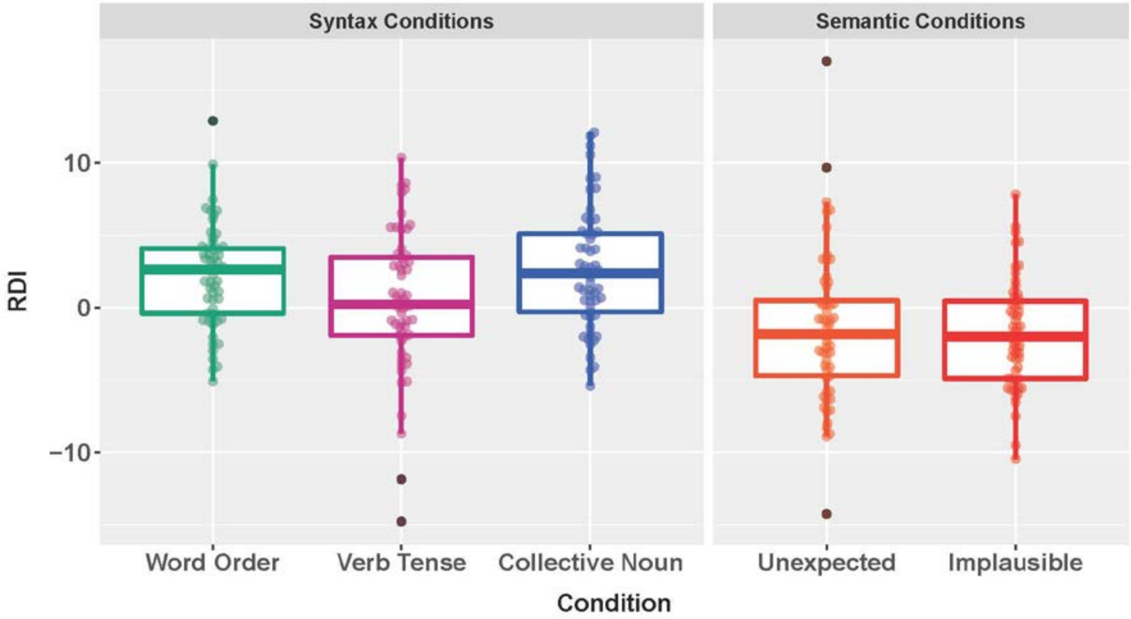

Figure 10. Boxplots overlaid with participant RDI values by condition in Experiment 1. Black points indicate boxplot outliers.

Pearson bivariate correlations computed between syntax RDI values and behavioral measures revealed a significant negative correlation between collective noun RDI values and reading experience scores, *r*(39) = -0.32, uncorrected *p* = 0.04. Individuals who were more N400-dominant in the collective noun condition had higher scores on the reading experience measure. There was an additional significant positive correlation between implausible RDI values and reading experience scores, *r*(39) = 0.35, uncorrected *p* = 0.027. In this case, participants who had higher reading experience scores were more P600-dominant when responding to implausible semantic continuations.

There were no significant correlations between RDI values and a-prime scores in any syntactic or semantic conditions (all *p*’s > .2). Participants were as sensitive to syntactic and semantic violations regardless of where on the RDI spectrum they fell. Furthermore, contrary to previous findings that showed response dominance patterns to syntactic violations are consistent within an individual, there were no significant correlations between RDI values in any of the syntactic conditions (all *p*'s > .3). Thus, an individual's response dominance to one type of syntactic violation did not predict whether that individual would be N400-dominant, P600-dominant, or biphasic in any other syntactic condition. A maximum likelihood confirmatory factor analysis was also not significant, $\chi^2(3, N = 57) = 1.41$, $p = 0.70$. RDI values across syntactic conditions did not appear to reflect an underlying factor related to syntactic processing.

3.1.2.1 Is variability in syntactic processing related to variability in language analytic ability?

Analyses revealed no relationship between variability in language analytic ability and syntactic processing. A linear model regressing language analytic ability on word order, verb tense, and collective noun RDI values was not significant, $F(3,53) = 0.41$, $p = 0.745$. A follow-up set of analyses correlated language analytic ability on RDI values in the three syntax conditions separately. No correlations were significant (all *p*'s > .4). Contrary to the predictions of the Syntactic Sensitivity account, variability in syntactic processing was not explained by differences in language analytic ability.

3.1.2.2 Does variability in syntactic processing reflect variability in the nature of individual's linguistic predictions in language processing?

Analyses comparing features of prediction across syntactic and semantic conditions revealed some relationships in type and strength of predictions across conditions. With regard to type of prediction, Table 7 contains the results of the linear models regressing implausible and unexpected RDI on word order, verb tense, and collective noun RDI. Together, word order, verb tense, and collective noun RDI explained 10% of the variance in implausible RDI scores $F(3,53)$

= 3.02, $p = 0.038$, $R^2 = 0.10$) and 13% of the variance in unexpected RDI scores ($F(3,53) = 3.57$, $p = 0.021$, $R^2 = 0.13$). However, only verb tense RDI was a significant predictor of RDI scores in the semantic conditions. Higher RDI values in the verb tense condition predicted higher RDI values in both implausible and unexpected conditions.

Table 7. Linear regression results for implausible and unexpected RDI outcomes in Experiment 1. Significant predictors are bolded.

	Beta	Std. Error	t-value	p-value	
Outcome: *Implausible RDI*					
(Intercept)	-1.81	0.62	-2.92	0.005	**
Word Order RDI	0.09	0.14	0.67	0.506	
Verb Tense RDI	**0.26**	**0.10**	**2.60**	**0.012**	*****
Collective Noun RDI	-0.11	0.11	-1.00	0.321	
Outcome: *Unexpected RDI*					
(Intercept)	-1.61	0.85	-1.91	0.062	†
Word Order RDI	0.10	0.19	0.55	0.583	
Verb Tense RDI	**0.41**	**0.13**	**3.05**	**0.004**	******
Collective Noun RDI	-0.09	0.16	-0.54	0.590	

Note. † $p < .01$; ** $p < .01$; *** $p < .001$.

Regression diagnostics revealed that data from five participants had high leverage on the model results, so follow-up bisquare robust linear regressions were computed. Results were qualitatively similar, and further revealed a marginal effect of word order RDI on variability in semantic processing (see Table 8).

Table 8. Bisquare robust linear regression results for implausible and unexpected RDI outcomes in Experiment 1. Significant predictors are bolded.

	Beta	Std. Error	t-value	p-value	
Outcome: *Implausible RDI*					
(Intercept)	-1.81	0.62	-2.92	< 0.001	***

Word Order RDI	0.09	0.14	0.67	0.083	†
Verb Tense RDI	**0.26**	**0.10**	**2.60**	**0.003**	**
Collective Noun RDI	-0.11	0.11	-1.00	0.629	
Outcome: *Unexpected RDI*					
(Intercept)	-2.52	0.66	-3.83	< 0.001	***
Word Order RDI	0.28	0.14	1.95	0.057	†
Verb Tense RDI	**0.46**	**0.10**	**4.43**	**< 0.001**	**
Collective Noun RDI	-0.14	0.12	-1.19	0.234	

Note. † $p < .01$; ** $p < .01$; *** $p < .001$.

Results indicated that individuals' response dominance patterns were similar across verb tense and semantic conditions, such that individuals who were more N400- or P600-dominant in response to verb tense violations were also more N400- or P600-dominant in response to implausible and unexpected continuations. To test whether individuals who were P600-dominant in syntax conditions also showed smaller N400 magnitudes in semantic conditions, individuals were separated according to their verb tense RDI values into N400-dominant (RDI < 0) and P600-dominant (RDI > 0) groups. ERPs to semantic conditions were plotted separately for the two groups and visual inspection suggested that both N400 and P600 magnitudes differed between conditions (see Figure 11).

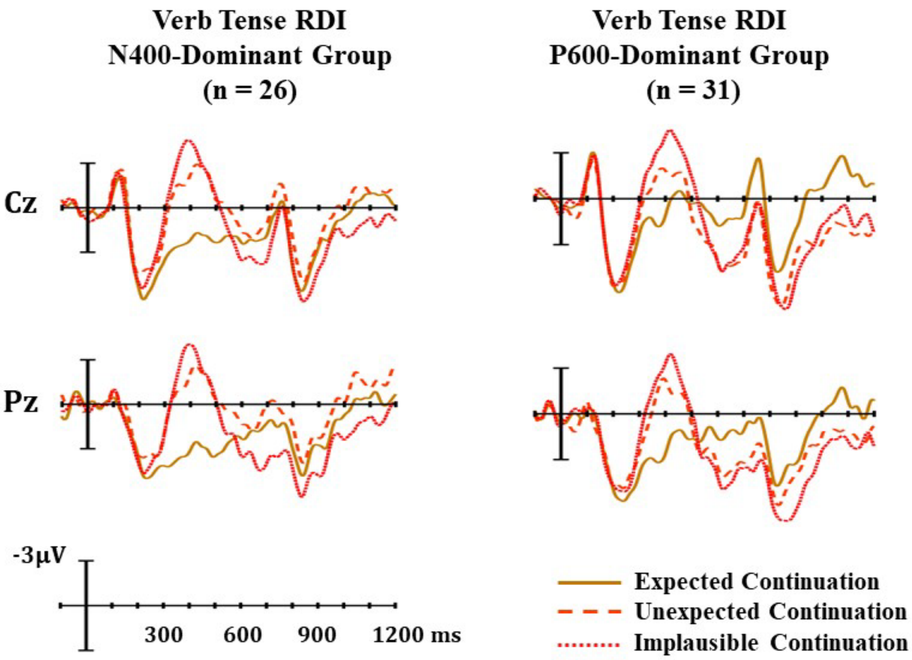

Figure 11. Grand-mean ERPs in semantic conditions by verb tense RDI dominance group in Experiment 1. Representative midline electrodes Cz and Pz from the ROI corresponding to the RDI are plotted.

One-tailed independent samples t-tests confirmed that the difference in N400 magnitude between conditions within the ROI corresponding to the RDI was significantly different in implausible conditions and marginal in unexpected conditions (see Table 9). Additionally, the difference in P600 magnitude between groups was marginal in implausible conditions and significant in unexpected conditions. Thus, individuals who were N400-dominant in response to verb tense errors were also showed larger N400 magnitudes in implausible conditions. In addition, unexpected continuations did not elicit P600 effects in these N400-dominant individuals compared to individuals who were P600-dominant in the verb tense condition.

Table 9. Means, standard deviations, and t- and p-values for differences in semantic condition ERP magnitudes for N400- and P600-dominant verb tense groups in

Experiment 1.

	Verb Tense RDI N400-Dominant Group (*n* = 25)	**Verb Tense RDI P600-Dominant Group (*n* = 31)**	***t*-value**	***p*-value**	
N400 Magnitude					
Implausible	4.92 (0.13)	3.45 (0.09)	-1.77	0.042	*
Unexpected	3.84 (0.17)	2.21 (0.11)	-1.53	0.066	†
P600 Magnitude					
Implausible	0.88 (0.13)	2.17 (0.11)	1.43	0.079	†
Unexpected	-0.71 (0.17)	2.28 (0.10)	3.01	0.002	**

Note. † *p* < .01; * *p* < .05; ** *p* < .01.

Results of models comparing strength of predictions across syntactic and semantic conditions suggested that there might be a relationship between beta desynchronization across syntactic conditions. Models regressing implausible and unexpected beta desynchronization on word order, verb tense, and collective noun beta desynchronization were not significant (*p*’s > .1). However, in the model of implausible beta desynchronization, word order beta desynchronization came out as a significant predictor (see Table 10). A follow-up bisquare robust linear regression confirmed that word order beta was a significant predictor of implausible beta desynchronization. The negative beta value indicated that the more beta desynchronized in response to word order violations, the less beta desynchronized in implausible violations.

Table 10. Linear and bisquare robust linear regression model results for implausible beta desynchronization in Experiment 1.

Outcome: *Implausible Beta*	Beta	Std. Error	*t*-value	*p*-value	
Linear Regression					
(Intercept)	0.26	0.07	3.65	0.001	***
Word Order Beta	**-0.23**	**0.11**	**-2.22**	**0.030**	*****
Verb Tense Beta	0.13	0.11	1.23	0.224	
Collective Noun Beta	-0.05	0.11	-0.46	0.647	

Robust Linear Regression					
(Intercept)	0.26	0.07	3.52	< 0.001	***
Word Order Beta	**-0.27**	**0.11**	**-2.45**	**0.018**	*
Verb Tense Beta	0.15	0.11	1.37	0.174	
Collective Noun Beta	-0.04	0.11	-0.34	0.735	

Note. * $p < .05$; *** $p \leq .001$.

3.1.2.3 Does variability in type of linguistic predictions made during syntactic processing relate to variability in strength of linguistic predictions?

Analyses testing whether variability in syntax RDI values could be explained by variability in beta desynchronization revealed that RDI values and beta desynchronization were correlated in the word order condition, $r(57) = 0.27$, $p = 0.045$, but not in the other syntax conditions (p's > .7). The positive correlation indicates that the degree of beta desynchronization in response to a word order violation was related to more P600-dominant RDI values in that condition. However, when English grammar proficiency, reading experience, and working memory scores were added as predictors of word order RDI values, the model was no longer significant ($F(4,33) = 1.46$, $p = 0.235$), although the effect of word order beta desynchronization was marginal ($t = 1.77$, $p = 0.085$). The strength of the correlation was weakened by the added predictors and the loss of power resulting from fewer available English grammar proficiency and reading experience scores.

Finally, analyses testing whether variability in syntax RDI values were related to degree of beta desynchronization in semantic conditions revealed no such relationship. Linear models regressing implausible and unexpected beta desynchronization on word order, verb tense, and collective noun RDI values were not significant (p's > .3). Thus, whether individuals were N400- or P600-dominant in response to syntactic errors did not predict the degree of beta desynchronization in response to semantic errors.

3.2 Experiment 2

3.2.1 *Whole-group results*

3.2.1.1 Behavioral results

There was a good deal of variability in performance on behavioral tasks across participants. Table 11 contains descriptive statistics for all predictor variables, including behavioral tasks measuring language analytic ability, working memory, and grammar proficiency, as well as and total years of English experience. Figure 12 shows the distribution of predictor variables. English grammar proficiency and total years of English experience were significantly positively correlated with each other, $r(48) = 0.32$, uncorrected $p = 0.024$. The more years of experience individuals had studying English and living in the U.S., the higher their English grammar proficiency scores were. In addition, there were marginal negative correlations between working memory scores and English grammar proficiency ($r(48) = -0.28$, uncorrected $p = 0.052$) and total years of English experience ($r(48) = -0.26$, uncorrected $p = 0.066$). The more experience Chinese speakers had with English and the better their English grammar proficiency, the worse they performed on the working memory task.

Table 11. Means, standard deviations, and ranges of predictor variables in Experiment 2.

	Mean	**SD**	**Range**
Language Analytic Ability	68.2	20.67	10-100
Working Memory	31.12	13.45	9-63
Grammar Proficiency	0.89	0.08	0.72-1
Total Years of English Experience	10.21	3.34	3-18

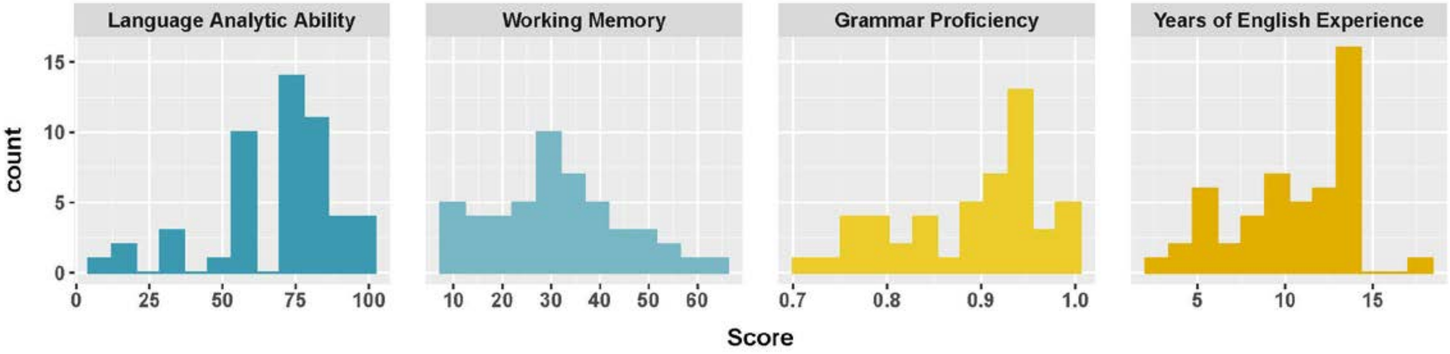

Figure 12. Distributions of scores of predictor variables in Experiment 2.

Chinese speakers generally had high a-prime scores across both Chinese and English stimuli but did appear to have lower sensitivity and greater variability in English verb tense scores (see Figure 13). A repeated- measures ANOVA indicated there was a significant difference in a-prime scores across conditions, $F(1.1, 54.07) = 44.67$, $p < .001$. FDR-corrected pairwise t-tests showed that Chinese word order a-prime (M = 0.98, SD = 0.02) was significantly higher than English word order a-prime (M = 0.94, SD = 0.03) and English verb tense a-prime (M = 0.86, SD = 0.11), which were also significantly different from each other.

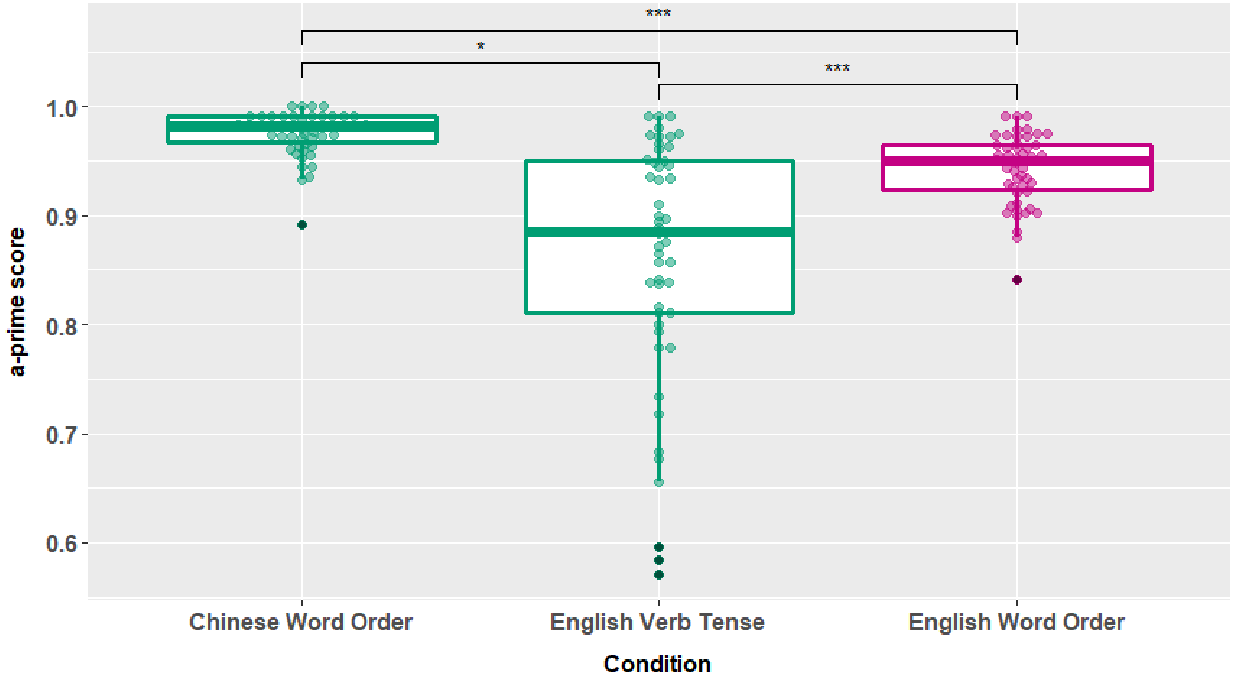

Figure 13. Boxplots of a-prime scores by condition in Experiment 2.

3.2.1.2 ERP results

3.2.1.2.1 Chinese condition

Grand mean waveforms for the Chinese word order condition are shown in Figure 14. Visual inspection of the waveforms indicated the presence of a widespread N400 effect elicited maximally in fronto-central regions and a P600 effect elicited maximally in the centro-parietal regions.

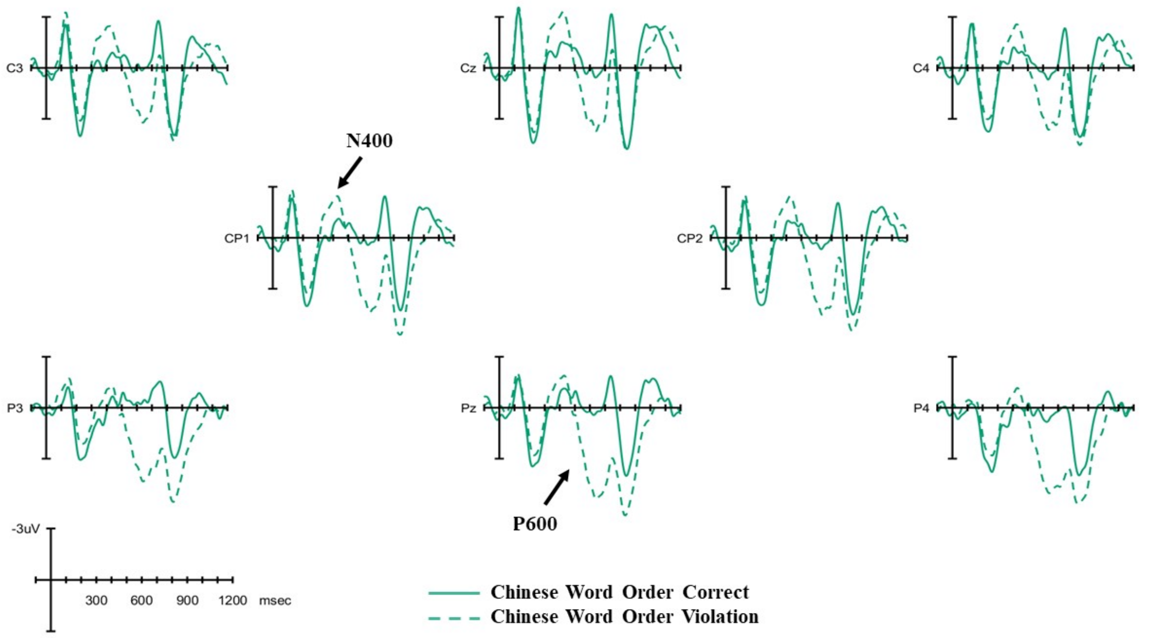

Figure 14. Grand mean ERP waveforms for the Chinese word order condition in Experiment 2. The mean activity for the grammatical condition is indicated by the solid line and the dashed line indicates mean activity for the ungrammatical condition.

Table 12 reports the results of the omnibus ANOVAs for Chinese word order. A main effect of grammaticality in the N400 window in midline and lateral electrode sites confirmed the presence of an N400 effect in the Chinese word order condition. N400 magnitude was more negative in the ungrammatical condition(Midline: M = -1.93 µV, SE = 0.18; Lateral: M = -1.29 µV, SE = 0.14) compared to the grammatical condition (M = -0.92 µV, SE = 0.2; Lateral: M = -0.26 µV, SE = 0.13).

Table 12. F-statistics from ANOVA results for midline and lateral electrodes in N400 and P600 windows for the Chinese word order condition in Experiment 1. Degrees of freedom are noted next to the main effects and interactions. Significant results are bolded.

	N400 (300-500 ms)	**P600** (500-800 ms)
Midline		
Gram. (1, 49)	**8.97****	**26.77*****
Ant. (2, 98)	**16.66*****	**38.9*****
Gram. x Ant. (2, 98)	3.27	**34.14*****
Lateral		
Gram. (1, 49)	**15.88*****	**28.08*****
Hem. (1, 49)	0.93	**25.64*****
Ant. (1, 49)	**51.42*****	**46.65*****
Gram. x Hem. (1, 49)	0.04	0.26
Gram. x Ant. (1, 49)	0.68	**40*****
Hem. x Ant. (1, 49)	1.58	**12.14*****
Gram. x Hem. x Ant. (1, 49)	1.02	2.23

Note. Gram. = Grammaticality; Hem. = Hemisphere; Ant. = Anteriority.

**$p < .01$, ***$p < 0.001$.

A main effect of grammaticality in the P600 electrodes in midline and lateral electrodes confirmed the presence of a P600 effect in the Chinese word order condition. P600 magnitude was more positive in the ungrammatical condition (Midline: M = 1.92 µV, SE = 0.28; Lateral: M = 1.48 µV, SE = 0.19) compared to the grammatical condition (M = -0.38 µV, SE = 0.19; Lateral: M = -0.32 µV, SE = 0.13). There was also a significant interaction between grammaticality and anterior, which indicated that the P600 effect was larger in posterior electrodes (Midline: M = 4 µV, SE = 0.56; Lateral: M = 2.79 µV, SE = 0.3) compared to anterior electrodes (Midline: M = 0.54 µV, SE = 0.47; Lateral: M = 0.81 µV, SE = 0.26). Chinese word order violations thus elicited widespread N400 effects and posteriorly-maximal P600 effects in

Chinese speakers.

3.2.1.2.2 English conditions

Grand mean waveforms for the English conditions are shown in Figure 15 and Figure 16. Visual inspection of the waveforms suggested the presence of a P600 effect in the English word order condition elicited maximally in posterior regions. In the English verb tense condition, there appeared to be a small N400 effect elicited maximally in left frontal electrodes and a P600 effect in central and right posterior electrodes.

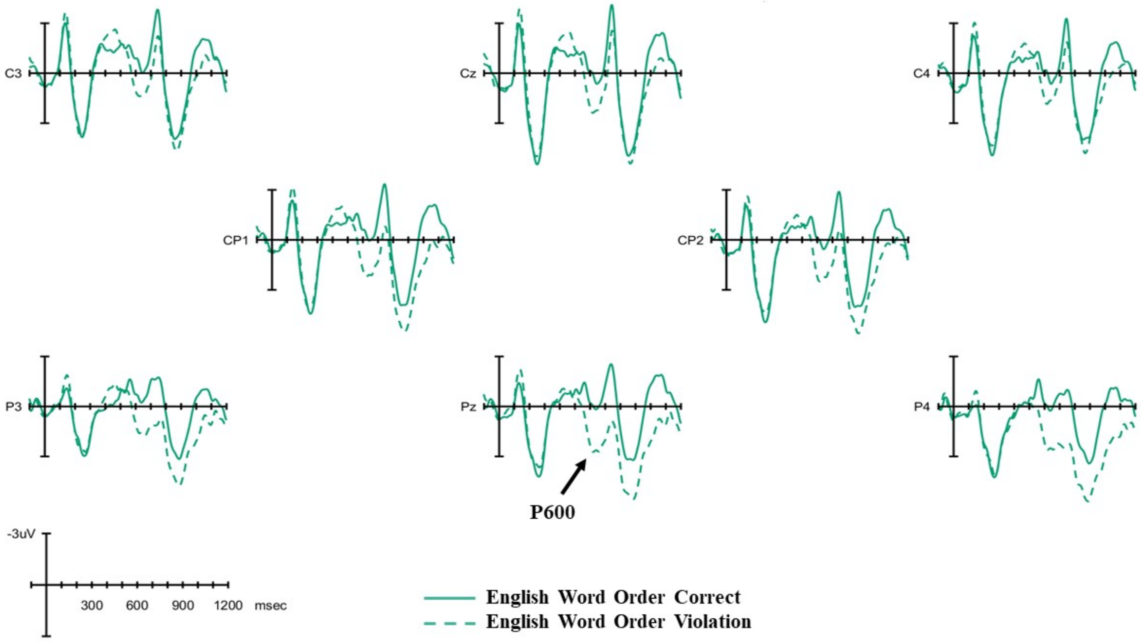

Figure 15. Grand mean ERP waveforms for the English word order condition in Experiment 2. The mean activity for the grammatical condition is indicated by the solid line and the dashed line indicates mean activity for the ungrammatical condition.

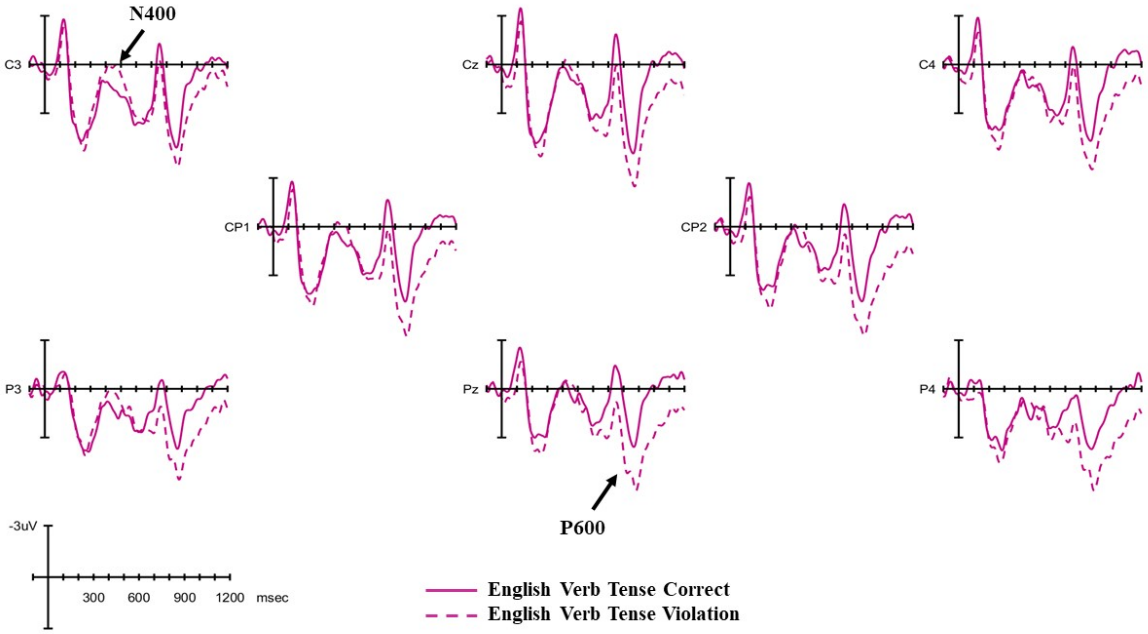

Figure 16. Grand mean ERP waveforms for the English verb tense condition in Experiment 2. The mean activity for the grammatical condition is indicated by the solid line and the dashed line indicates mean activity for the ungrammatical condition.

Table 13 reports the results of the omnibus ANOVAs for English stimuli. There was no main effect of or interactions with grammaticality in the N400 window in midline or lateral electrode sites. However, there was a grammaticality by hemisphere interaction in lateral electrodes, showing that the difference between grammatical and ungrammatical conditions was larger in the left hemisphere (M = 0.61 µV, SE = 0.2) compared to the right (M = -0.11 µV, SE = 0.21). This interaction was qualified by a significant interaction with condition. Follow-up ANOVAs showed that the grammaticality by condition interaction was significant in the left hemisphere ($F(1, 99) = 4.93$, $p = 0.029$) but not in the right ($F(1, 99) = 0.98$, $p = 0.325$). The difference in magnitude between grammatical and ungrammatical conditions was larger in the left hemisphere in the English verb tense condition (M = 0.95 µV, SE = 0.31) compared to the

English word order condition (M = 0.26 µV, SE = 0.25). The grammaticality by condition by hemisphere interaction was further qualified by an interaction with anteriority. Follow-up ANOVAs showed that the grammaticality by condition interaction was significant in left frontal electrodes (*F*(1, 49) = 10.62, *p* = 0.002), but not in left posterior (*F*(1, 49) = 0.09, *p* = 0.764), right frontal (*F*(1, 49) = 1.54, *p* = 0.22), or right posterior electrodes (*F*(1, 49) = 0.03, *p* = 0.859). The difference between grammatical and ungrammatical conditions was larger in left frontal electrodes in the English verb tense condition (M = 1.28 µV, SE = 0.41) than in the English word order condition (M = 0.05 µV, SE = 0.33). Thus, the omnibus ANOVA confirmed the presence of an N400 effect in left frontal electrodes in the English verb tense condition.

Table 13. F-statistics from ANOVA results for midline and lateral electrodes in N400 and P600 windows for English conditions in Experiment 2. Degrees of freedom are noted next to the main effects and interactions. Significant results are bolded.

	N400 (300-500 ms)	**P600** (500-800 ms)
Midline		
Gram. (1, 49)	0.67	**6.67***
Cond. (1, 49)	**24.58*****	**32.95*****
Ant. (2, 98)	0.35	0.68
Gram. x Cond. (1, 49)	0.06	0.02
Gram. x Ant. (2, 98)	1.47	**12.63*****
Cond. x Ant. (2, 98)	**22.71*****	**21.28*****
Gram. x Cond. x Ant. (2, 98)	0.25	**6.84****
Lateral		
Gram. (1, 49)	0.62	**5.63***
Cond. (1, 49)	**52.98*****	**61.08*****
Hem. (1, 49)	**19.42*****	**9.95****
Ant. (1, 49)	**23.66*****	1.52
Gram. x Cond. (1, 49)	0.23	1.52
Gram. x Hem. (1, 49)	**23.3*****	**11.93*****
Cond. x Hem. (1, 49)	**8.18****	**4.77***

Gram. x Ant. (1, 49)	0.04	**6.56***
Cond. x Ant. (1, 49)	**32.27*****	**27.37*****
Hem. x Ant. (1, 49)	**6.61***	**5.1***
Gram. x Cond. x Hem. (1, 49)	**12.09*****	**11.3****
Gram. x Cond. x Ant. (1, 49)	0.92	3.6
Gram. x Hem. x Ant. (1, 49)	0.42	0.24
Cond. x Hem. x Ant. (1, 49)	0.17	3.59
Gram. x Cond. x Hem. x Ant. (1, 49)	**14.93*****	**16.92*****

Note. Gram. = Grammaticality; Cond. = Condition; Hem. = Hemisphere; Ant. = Anteriority. *$p < 0.05$; **$p < .01$; ***$p < 0.001$.

In the P600 window, there was a widely distributed main effect of grammaticality across midline and lateral electrodes, such that ungrammatical conditions had larger P600 magnitudes (Midline: M = 1.59 µV, SE = 0.21; Lateral: M = 1.34 µV, SE = 0.15) relative to grammatical conditions (Midline: M = 0.42 µV, SE = 0.17; Lateral: M = 0.45 µV, SE = 0.13). The magnitude of the P600 effect across conditions was larger in posterior electrodes. A grammaticality by anteriority interaction in midline and lateral electrodes indicated that the P600 effect was larger in posterior electrodes (Midline: M = 1.19 µV, SE = 0.41; Lateral: M = 1.17 µV, SE = 0.24) compared to anterior electrode (Midline: M = 1.19 µV, SE = 0.41; Lateral: M = 0.6 µV, SE = 0.26). In midline electrodes, there was also a grammaticality by condition by anteriority interaction. Follow-up ANOVAs showed that the grammaticality by anteriority interaction was significant in the English word order condition ($F(1.47, 72.11) = 18.44$, $p < .001$) but not the English verb tense condition ($F(1.48, 72.42) = 0.72$, $p = 0.453$). The P600 effect was thus larger in the English word order condition in posterior electrodes (M = 2.34 µV, SE = 0.58) compared to anterior electrodes (M = 0.09 µV, SE = 0.47). In lateral electrodes, there was a significant grammaticality by hemisphere interaction that was also qualified by an interaction with condition, and follow-up ANOVAs showed the grammaticality by condition interaction was significant in left hemisphere electrodes (*F*(1, 99) = 11.02, *p* = 0.001) but not in right hemisphere electrodes (*F*(1, 99) = 0.06, *p* = 0.809). In the left hemisphere, the P600 effect was larger in the English word order condition (M = 1.16 µV, SE = 0.31) compared to the English verb tense condition (M = -0.06 µV, SE = 0.36). Finally, there was a significant four-way grammaticality by condition by hemisphere by anteriority interaction in lateral electrodes. Follow-up ANOVAs showed that the grammaticality by condition interaction was significant in left frontal electrodes (*F*(1, 49) = 7.52, *p* = 0.008), but not in left posterior (*F*(1, 49) = 3.8, *p* = 0.057), right frontal (*F*(1, 49) = 3.1, *p* = 0.085), or right posterior electrodes (*F*(1, 49) = 2.94, *p* = 0.093). The difference between grammatical and ungrammatical conditions was larger in left frontal electrodes in the English word order condition (M = 0.92 µV, SE = 0.39) than in the English verb tense condition (M = -0.48 µV, SE = 0.49). Thus, ANOVA results confirmed that both conditions showed a P600 effect but the P600 effect was larger and more widely distributed in the English word order condition compared to the English verb tense condition.

3.2.1.3 Time-frequency results

Mean beta power by grammaticality and condition in the syntax conditions are plotted in Figure 17. A two-way repeated-measures ANOVA analyzing differences in beta desynchronization across grammaticality and Chinese and English conditions revealed a main effect of grammaticality (*F*(1, 49) = 82.84, *p* < .001) and a main effect of condition (*F*(2, 98) = 6.48, *p* = 0.002), but no interaction between grammaticality and syntax condition (*p* = 0.293). The beta power was smaller in ungrammatical conditions (M = -0.51 dB, SE = 0.04) relative to grammatical conditions (M = -0.12 dB, SE = 0.04). FDR-corrected pairwise t-tests showed that the English verb tense condition had lower beta power across both grammaticality conditions (M = -0.45 dB, SE = 0.06) compared to the Chinese word order (M = -0.29 dB, SE = 0.04) and

English word order conditions (M = -0.19 dB, SE = 0.05), which did not differ from one another. Overall, ungrammatical conditions resulted in greater beta desynchronization relative to grammatical conditions, and the degree of beta desynchronization did not differ across syntactic conditions. This replicates the finding that syntactic violations result in smaller beta values compared to syntactically well-formed stimuli and extends this finding to the context of L2 language processing. Thus, in all Chinese and English conditions, violations interrupted the maintenance of the top-down cognitive set of the sentence meaning, and the degree of this interruption was similar across languages and rules.

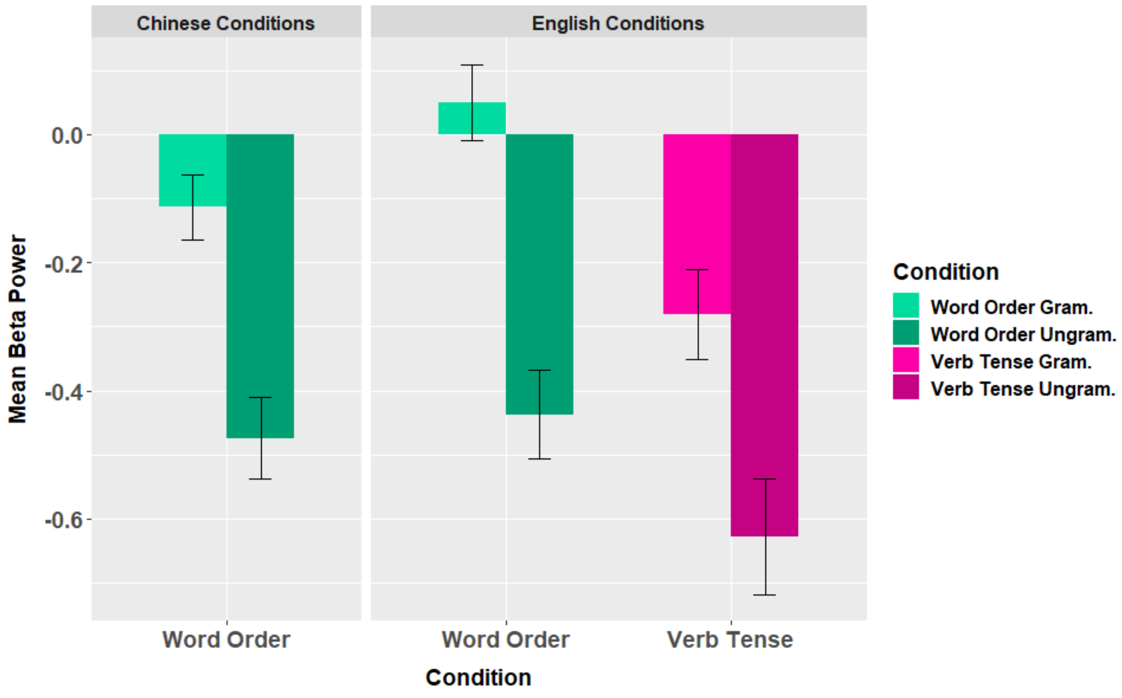

Figure 17. Mean beta power by grammaticality and condition in Experiment 2. Error bars indicate one standard error above and below the mean.

3.2.2 *Individual differences results*

In Chinese and English conditions, there was a good deal of variability in participant response patterns. Figure 18 shows each participant’s N400 and P600 magnitudes plotted against each other for each condition. Similar to previous research, there was a significant negative relationship between N400 effect magnitude and P600 effect magnitude in Chinese word order (r(48) = -0.58, p < .001) and English word order conditions (r(48) = -0.71, p < .001). However, the correlation between N400 and P600 effect magnitude in the English verb tense condition was positive, r(48) = 0.73, p < .001. Chinese learners did not generally show dominance of N400 or P600 effects in this condition, clustering around the equal effect magnitude line. Inspection of individual waveforms indicated that some learners had not yet started showing either ERP effect, or that the difference between grammatical and ungrammatical verb tense conditions was similar across the two time windows. For instance, some participants showed a prolonged N400 effect that spanned a large portion of the 1200 ms epoch. These data indicate that as a group, learner’s neural responses to English verb tense conditions were still in development, although electrophysiological behavior varied across learners. That is, in line with previously observed differences among learners, some individuals had already acquired an ERP response dominance pattern while others had not yet begun to discriminate grammatical and ungrammatical uses of the English verb tense rule.

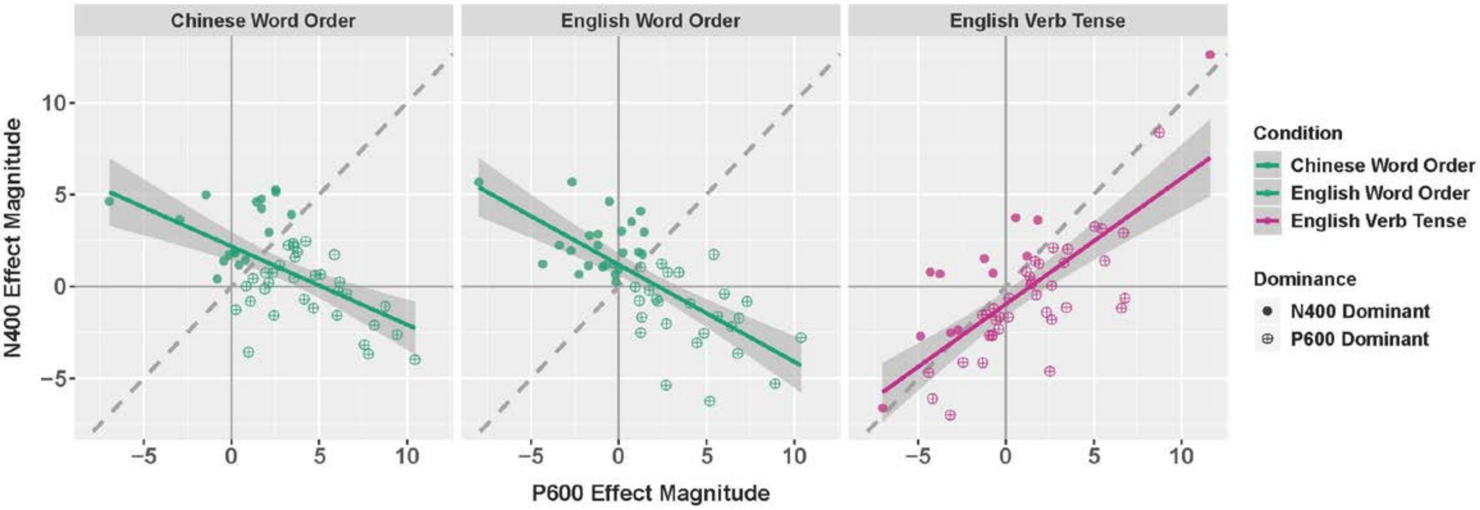

Figure 18. Participant N400 and P600 effect magnitudes by condition in Experiment 2. Dotted grey lines indicate line of equal ERP effect size. Solid circles to the left of the equal dominance line indicate N400-dominant individuals and crossed open circles the right of the equal dominance line indicate P600-dominant individuals. Solid color lines indicate regression lines.

Chinese speakers appeared to be similar in their response dominance patterns in the word order condition across languages. Although many participants appeared to be P600-dominant in the English verb tense condition, many of those individuals remained close to the line of equal effect magnitudes, suggesting that the P600 effect in these learners was not overwhelmingly larger than the N400 effect. Learners in the English verb tense condition also appeared to be less variable in their RDI values than in Chinese and English word order. Table 14 shows a summary of mean RDI values and proportion of P600-dominant individuals by condition. Figure 19 provides boxplots of RDI values by condition with individual RDI values overlaid. A repeated-measures ANOVA revealed that RDI values across conditions were not significantly different from each other, $F(2, 98) = 0.57$, $p = 0.567$. Mauchly's test of sphericity was also not significant ($p = 0.096$). RDI values in Chinese and English conditions had similar degrees of variability.

Nevertheless, the pattern of responses in the English verb tense condition suggests individuals' neural responses were qualitatively different for this unique L2 rule compared to the familiar rule in L1 and L2, even if the means and variances in RDI values were not statistically different.

Table 14. Means and standard deviations of RDI values and proportion of P600-dominant individuals by condition in Experiment 2.

	RDI Mean (SD)	**Proportion P600-Dominant**
Chinese Conditions		
Word Order	1.46 (3.55)	0.58
English Conditions		
Word Order	0.95 (4.04)	0.46
Verb Tense	0.87 (1.85)	0.67

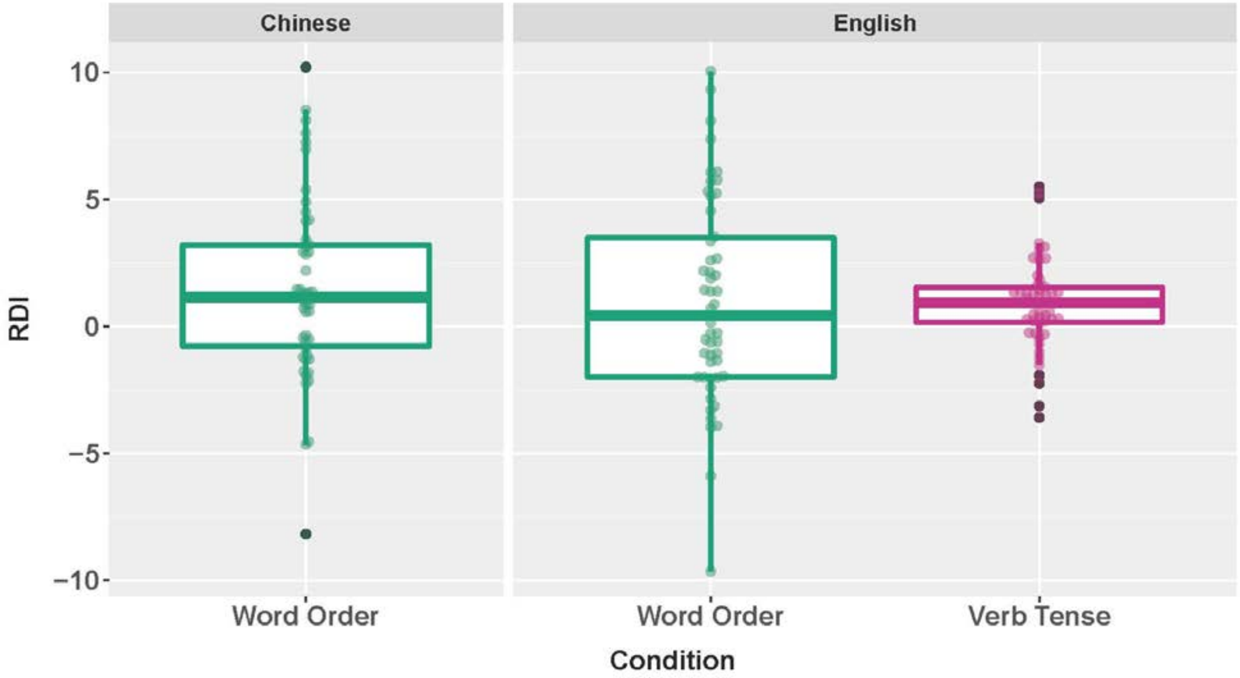

Figure 19. Boxplots overlaid with participant RDI values by condition in Experiment 2. Black points indicate boxplot outliers.

Pearson bivariate correlations computed between RDI values and behavioral measures revealed a significant positive correlation between English grammar proficiency and English word order RDI (*r*(48) = 0.36, uncorrected *p* = 0.011), as well as English verb tense RDI (*r*(48) = 0.33, uncorrected *p* = 0.02). Higher scores on the English grammar proficiency test were associated with larger RDI values. The more proficient learners were, the more P600-dominant their ERP response patterns were.

RDI values and a-prime scores were positively correlated in the Chinese word order condition (*r*(48) = 0.39, *p* = 0.006) and in the English verb tense condition (*r*(48) = 0.36, *p* = 0.011), but not in the English word order condition (*p* = 0.986). In both Chinese word order and English verb tense conditions, greater sensitivity to syntactic violations was related to larger RDI values (more P600-dominant responses). As for the English word order condition, participants were as sensitive to violations regardless of where on the RDI spectrum they fell.

3.2.2.1 Is variability in L2 unique rule processing related to variability in language analytic ability?

Analyses revealed no relationship between variability in language analytic ability and L2 unique rule processing. When language analytic ability, English grammar proficiency, and total years of English experience were entered as predictors of English verb tense RDI, the model was not significant, *F*(3,46) = 2.10, *p* = 0.113. The linear model of English verb tense a-prime was marginal, *F*(3,46) = 2.78, *p* = 0.052. However, only English grammar proficiency came out as a significant predictor (see Table 15). Contrary to the predictions of the Syntactic Sensitivity account, variability in L2 unique rule processing was not explained by differences in language analytic ability when controlling for L2 language experience.

Table 15. Linear regression results for the model testing the relationship between English verb tense a-prime and language analytic ability in Experiment 2. Significant predictors are bolded.

	Beta	Std. Error	*t*-value	*p*-value	
Outcome: *English Verb Tense a-prime*					
(Intercept)	0.35	0.18	2.00	0.052	†
Language Analytic Ability	0.00	0.00	-0.06	0.956	
English Grammar Proficiency	**0.60**	**0.22**	**2.76**	**0.008**	**
Total Years of English Experience	0.00	0.00	-0.42	0.674	

Note. † $p < .01$; ** $p < .01$.

3.2.2.2 Does variability in L2 unique rule processing reflect variability in the nature of individual's linguistic predictions in language processing?

Analyses comparing features of prediction across L2 unique rule and L1 conditions revealed contribution of type, but not strength, of L1 predictions. With regard to type of prediction, the linear model regressing English verb tense RDI on Chinese word order RDI, English word order RDI, English grammar proficiency, and total years of English experience was not significant, $F(4,45) = 1.85$, $p = 0.136$. However, the same model with English verb tense a-prime as the outcome variable was significant: together, English verb tense RDI on Chinese word order RDI, English word order RDI, English grammar proficiency, and total years of English experience explained 24% of the variance English verb tense a-prime scores $F(4,45) = 4.77$, $p = 0.003$, $R^2 = 0.24$ (see Table 16). Chinese word order RDI was a significant predictor of English verb tense a-prime scores. Higher RDI values (more P600-dominant responses) in the Chinese word order condition were associated with higher a-prime scores in the English verb tense condition. English word order RDI was also a significant predictor, but in this case smaller RDI values (more N400-dominant responses) predicted higher a-prime scores. This relationship only occurred after controlling for the other predictor variables as a Pearson bivariate correlation between English word order RDI and English verb tense a-prime was not significant (*p* = 0.581). Finally, English grammar proficiency was also a significant predictor of a-prime scores, such that higher scores on the grammar proficiency test were related to higher English verb tense a-prime scores.

Table 16. Linear regression results for the model testing the relationship between English verb tense a-prime and Chinese word order RDI in Experiment 2. Significant predictors are bolded.

	Beta	Std. Error	*t*-value	*p*-value	
Outcome: *English Verb Tense a-prime*					
(Intercept)	0.21	0.17	1.23	0.224	
Chinese Word Order RDI	**0.01**	**0.00**	**2.38**	**0.022**	*****
English Word Order RDI	**-0.01**	**0.00**	**-2.33**	**0.025**	*****
English Grammar Proficiency	**0.76**	**0.21**	**3.67**	**0.001**	*******
Total Years of English Experience	0.00	0.00	-0.70	0.490	

Note. * $p < .05$; *** $p < .001$.

Results indicated that individuals' L1 response dominance patterns predicted performance in the unique L2 rule, over and above L2 language experience. To test whether learners who had higher a-prime scores were also P600-dominant in their L1, participants were separated into high and low English verb tense a-prime groups according to a median split. ERPs to semantic conditions were plotted separately for the two groups and visual inspection suggested that both N400 and P600 magnitudes differed between groups in the Chinese word order condition (see Figure 20Figure 11).

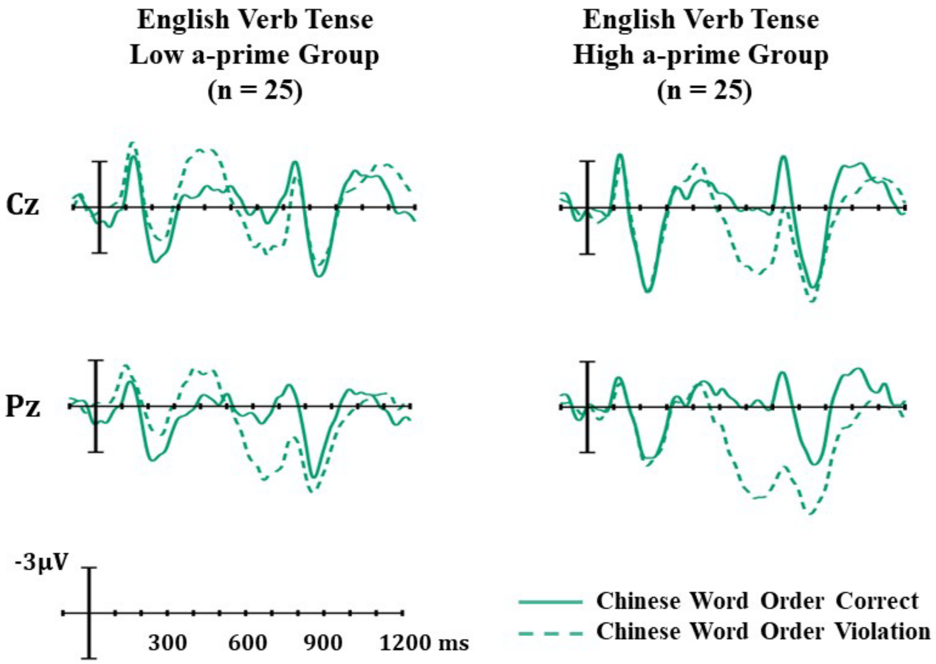

Figure 20. Grand-mean ERPs in Chinese word order by English verb tense a-prime group in Experiment 2. Representative midline electrodes Cz and Pz from the ROI corresponding to the RDI are plotted.

A one-tailed independent samples t-test confirmed that Chinese word order P600 magnitude was significantly larger in the high English verb tense a-prime group (M = 1.81, SE = 0.09) compared to the low English verb tense a-prime group (M = 0.01, SE = 0.09), $t(48) = 1.99$, $p = 0.026$. A one-tailed independent samples t-test also confirmed that Chinese word order N400 magnitude was significantly smaller in the high English verb tense a-prime group (M = 2.09, SE = 0.14) compared to the low English verb tense a-prime group (M = 3.57, SE = 0.11), $t(48) = -2.83$, $p = 0.003$. Thus, learners who were P600-dominant in their L1, showing smaller N400 magnitudes and larger P600 magnitudes to Chinese word order violations, were also more sensitive to errors in the L2 unique rule.

In analyses testing the influence of strength of prediction in L1 on the L2 unique rule, the linear model regressing English verb tense beta desynchronization on Chinese word order beta desynchronization, English word order beta desynchronization, English grammar proficiency, and total years of English experience was not significant, $F(4,45) = 0.36$, $p = 0.838$, $R^2 = -0.06$. The linear model of English verb tense a-prime, on the other hand, was significant, $F(4,45) = 2.97$, $p = 0.029$, $R^2 = 0.14$. However, only English grammar proficiency was a significant predictor (see Table 17). As in the case of the prediction type models, higher English grammar proficiency scores predicted higher English verb tense a-prime scores.

Table 17. Linear regression results for the model testing the relationship between English verb tense a-prime and Chinese word order beta desynchronization in Experiment 2. Significant predictors are bolded.

	Beta	Std. Error	*t*-value	*p*-value	
Outcome: *English Verb Tense a-prime*					
(Intercept)	0.25	0.19	1.31	0.195	
Chinese Word Order Beta	-0.07	0.04	-1.56	0.127	
English Word Order Beta	-0.03	0.04	-0.73	0.469	
English Grammar Proficiency	**0.65**	**0.21**	**3.03**	**0.004**	**
Total Years of English Experience	0.00	0.00	-0.01	0.989	

Note. * $p < .05$; *** $p < .001$.

3.2.2.3 Does variability in type of linguistic predictions made during L2 unique rule processing relate to variability in strength of linguistic predictions?

Variability in L2 unique rule processing was not found to be related to the strength of linguistic predictions learners made in the L2 unique rule or in L1. The model regressing English verb tense RDI on English verb tense beta desynchronization, English grammar proficiency, and years of English experience was not significant, $F(3,46) = 2.07$, $p = 0.117$, $R^2 = 0.06$. The model regressing English verb tense RDI on Chinese word order beta desynchronization, English word order beta desynchronization, English grammar proficiency, and years of English experience was also not significant, $F(4,45) = 1.62$, $p = 0.186$, $R^2 = 0.05$. Thus, whether learners were N400- or P600-dominant in response to L2 unique rule errors was not related to the degree of beta desynchronization in response to the L2 unique rule or in L1.

Chapter 4. DISCUSSION

The experiments described herein examined whether variability in neural responses to grammatical errors in L1 and L2 could be accounted for by individual differences in the way comprehenders build predictions of upcoming input or in comprehenders' sensitivity to grammatical form. The findings provide converging evidence that previously observed variability in L1 and L2 syntactic processing may arise from an individual's dynamic use of semantic versus syntactic information to drive predictions during comprehension.

Individual differences in syntactic processing predicted how comprehenders processed linguistic errors in other linguistic anomalies, in both native and non-native language processing contexts. Native English speakers who showed more P600-dominant ERP patterns to syntactic errors showed quantitatively and qualitatively different ERP responses than N400-dominant individuals when they encountered semantic anomalies: the magnitude of their N400 response to semantic anomalies was smaller and they additionally showed P600 effects in the unexpected condition. Native Chinese speakers who showed more P600-dominant ERP response patterns to syntactic errors in Chinese also showed greater proficiency in a complex L2 English grammatical rule than learners who were N400-dominant in Chinese. Thus, the syntactic processing differences observed in participants predicted variability in semantic processing and unfamiliar L2 grammatical rule acquisition, two conditions that demonstrate these individual differences are not just a feature of L1 grammatical comprehension.

This conclusion is further supported by the finding that the observed variability among participants could not be explained by differences in language analytic ability, indicating that whether an individual was N400- or P600-dominant in response to syntactic errors was not a reflection of that individual's sensitivity to grammatical form. The data are in fact incompatible with this explanation for the nature of individual differences in syntactic processing. First, if individual differences in syntactic processing reflected a sensitivity to grammatical form, then whether a comprehender was more or less sensitive to syntactic structure would have had no bearing on how that individual processed semantic anomalies where the syntactic structure was intact. Furthermore, if P600-dominant responses to L1 syntactic errors indicated better language analytic abilities, then L1 processing patterns could have facilitated acquisition of familiar grammatical rules in L2, but not unfamiliar grammatical rules for which a learner would struggle to develop analytical awareness.

Individual differences in syntactic processing instead appear to characterize an aspect of language processing that is recruited dynamically by speakers in a systematic way. Specifically, the data suggest that the difference between N400-dominant and P600-dominant individuals previously observed in grammatical anomaly processing may lie in whether a comprehender builds stronger semantic or syntactic predictions during language comprehension, respectively. Form-focused predictors build stronger syntactic expectations that are consequently strengthened or adjusted upon encountering the upcoming input. This permits form-focused individuals to acquire new complex grammatical rules more effectively than meaning-focused predictors as prediction error feedback from the input informs the knowledge base and improves subsequent predictions. Indeed, acquisition of a unique L2 grammatical rule was facilitated in learners whose L1 syntactic error processing patterns were characterized by greater violations of syntactic expectations. However, being a syntax-focused individual may also come at the cost of sensitivity to semantic nuance. When predictions are based on syntactic information, semantic expectations are not as strong. As a result, form-focused individuals were not perturbed by semantic anomalies to the same degree as meaning-focused comprehenders.

Investigation into the role of factors previously used to explain individual differences in language processing showed that working memory ability, native language proficiency, and language experience all failed to account for variability in these data. For instance, individual differences in syntactic processing predicted variability in semantic processing and complex L2 rule acquisition even when controlling for working memory. There were also no differences in working memory scores between N400-dominant and P600-dominant participants of either language group. Similarly, reading experience and grammar proficiency did not explain variability in semantic error processing over and above syntactic error processing patterns, and the role of L1 syntax processing variability in L2 unique rule acquisition remained even when controlling for neural and behavioral measures of L2 language experience. In sum, whether a comprehender built stronger semantic or syntactic predictions during language comprehension was not predicted by several cognitive or language experience variables. This demonstrates that individual differences in linguistic predictions are a unique feature of the language comprehension process that exists independently of a comprehender's strengths in other relevant attributes.

There was little evidence that individual differences in ERP processing patterns were related to variability in beta desynchronization, or that individuals differed systematically in the degree of beta desynchronization across linguistic anomalies. With the exception of a positive correlation between RDI values and beta desynchronization in the word order condition among English participants, there were no other relationships between the type of predictions comprehenders built and the strength of those predictions. This suggests that although individuals may differ in the type of predictions they build for comprehension, they do not appear to differ to the same degree in terms of the overall strength of their predictions. Thus, linguistic errors generally interrupted individuals' top-down contextual predictions to similar degrees regardless of whether those predictions were semantically or syntactically driven.

4.1 The status of the N400 in syntactic processing

4.1.1 *The N400 in L1 grammar comprehension*

Taken together, the findings in these experiments not only present evidence that individuals differ in the way they build linguistic predictions, but crucially underscore that neither form-focused nor meaning-focus individuals necessarily make for better comprehenders. These data suggest a need for reevaluating what it means for an individual to demonstrate N400-dominant ERP responses during grammar comprehension. Under the interpretation that the P600 reflects processes related to syntactic analysis and repair, the absence or attenuation of the P600 effect has been argued to signal impaired or underdeveloped grammar comprehension mechanisms (Canette et al., 2020; Clahsen et al., 2007; Hahne et al., 2004; Kielar et al., 2019; Pakulak & Neville, 2010). Studies have also related smaller P600 effects and larger N400 magnitudes with lower language proficiency scores in healthy adult native speakers (Pakulak & Neville, 2010; C. Weber-Fox et al., 2003). However, aligning with related research on individual differences in language processing (Mehravari et al., 2017; Tanner, 2019; Tanner & Van Hell, 2014), these data show that N400 responses to grammatical errors need not reflect atypical or impaired syntactic processing. In all native language syntax error processing contexts, some portion of English and Chinese participants were identified as being N400-dominant, indicating that this pattern of ERP responses is part of the natural variability found among native speakers. In fact, there were even instances when N400-dominant individuals outperformed P600-dominant individuals. For example, higher reading experience scores were related to N400-

dominant responses in the collective noun condition among English speakers, a result that is difficult to reconcile with the perspective that N400 responses occur in individuals who are struggling with grammatical processing. The collective noun condition violated agreement between a noun that refers to an entity composed of multiple elements and its subsequent main verb (e.g., *Later tonight, the committee *have an important meeting*). Within the context of the experiment, repeated encounters with the collective noun may have led more experienced readers to attend more closely to its semantic features and thus build stronger meaning-based predictions for the appropriate verb form by honing in on the singular entity of the noun. In this way, increased experience with reading enabled individuals to make greater use semantic information to inform predictions when it was advantageous to do so, encouraging an alternative but no less successful language processing strategy. Just as the highly skilled deaf readers in Mehravari et al. (2017) were found to achieve efficient written language comprehension by attending to different linguistic cues than highly skill hearing readers, N400-dominant participants in these experiments arrived at similarly effective comprehension as P600-dominant participants did despite taking a different path.

Contrary to previous work (Pakulak & Neville, 2010), there was also no evidence that P600-dominant individuals were more proficient in their native language than N400-dominant individuals. In both English and Chinese participants, grammar proficiency scores did not differ based on whether an individual showed P600- or N400-dominant ERP response patterns when processing syntactic errors in their native language. In fact, individual difference in verb tense processing remained a significant predictor of variability in semantic anomaly processing even when controlling for grammar proficiency or overall proficiency, where scores ranged from 80-99% for both measures (SDs = .04). Although it is possible the proficiency tests were not as sensitive as those used by Pakulak and Neville (2010), who found N400 effects in low-proficiency, low SES groups, it is notable that N400-dominant ERP patterns were observed even among highly educated and skilled speakers in the current experiment. There was also no relationship between RDI values in any of the syntactic or semantic conditions and a-prime scores, demonstrating that engaging with semantic information did not preclude individuals from achieving the same levels of comprehension as their peers who engaged instead with syntactic cues. In short, N400-dominant ERP patterns in response to grammatical anomalies should not in themselves be considered a marker of poor grammatical or language proficiency.

It thus appears that comprehenders may attend to semantic cues or syntactic cues to build predictions, and that both approaches result in successful and efficient comprehension from two different directions. If P600 effects reflect violations of syntactic predictions, then individuals who show more P600-dominant responses to syntactic errors can be understood to focus more on building syntactic predictions to ease comprehension. This processing strategy leads syntactically driven individuals to prioritize syntactic cues. As a result, semantic anomalies elicit smaller N400 magnitudes because their semantic predictions are not as strong as those of semantically driven individuals. N400-dominant individuals, on the other hand, prioritize semantic cues and build semantic predictions to ease comprehension. This makes them more aware of semantic information that deviates from meanings supported by the context. In this way, N400-dominant individuals may also be well practiced in quickly extracting meaning and adapting their representation during reading, a conclusion further supported by the relationship between N400-dominance in the collective noun condition and reading experience scores. Consequently, it is not clear that N400-dominant comprehenders are engaging in grammatical processing that is simply “good-enough” (Ferreira, 2003; Ferreira et al., 2002; Ferreira & Patson, 2007). These individuals build rich representations of meaning even from fairly sparse sentence contexts and are capable of tackling the syntactic violations as successfully as grammatically driven individuals.

4.1.2 *The N400 in L2 grammar learning*

The data from Experiment 2 add to the large body of evidence that as learners increase in their L2 rule sensitivity and L2 proficiency, larger P600 effects are elicited in response to syntactic violations (Kotz, 2009; Rossi et al., 2006b; Tanner et al., 2013; van Hell & Tokowicz, 2010; White et al., 2012). In Chinese speakers, grammar proficiency scores were positively correlated with RDI values in both English rules. Although the present study did not track acquisition longitudinally, it did appear that P600 effects take longer to develop for unique L2 rules and that P600 effects occur in individuals who have higher sensitivity to the rule. On the one hand, this is in line with evidence that as learners acquire proficiency in a new language, they appear to transition to showing P600 effects in response to syntactic errors (McLaughlin et al., 2010; Osterhout et al., 2006, 2008; Tanner et al., 2013). However, data from English speakers in Experiment 1 showed that even native speakers can be N400-dominant, so it is possible that some N400-dominant learners also had high sensitivity to the L2 unique rule. In fact, several individuals with N400-dominant responses to English verb tense were found to have a-prime scores at or above the mean of individuals with P600-dominant responses. Thus, these data put the status of the P600 as the definitive marker of "end-state" learning into question. This is in line with findings that English learners of Spanish as well as early and late Spanish-English bilinguals show either N400- or P600-dominant ERP patterns to morphosyntactic violations across the two languages (Bice & Kroll, 2017; Finestrat et al., 2020). Not only are N400-

dominant responses part of the natural variability in L1 processing, but that variability transfers to and occurs in L2 processing contexts, regardless of whether the L2 was acquired early in life or later in adulthood.

Moreover, semantically driven processing may sometimes be advantageous for L2 acquisition. Although higher a-prime and grammar proficiency scores were generally related to more positive RDI values (more P600-dominant responses), there was also evidence that smaller RDI values (more N400-dominant responses) were related to better performance in the familiar L2 rule. English word order RDI was negatively correlated with English verb tense a-prime scores, indicating that individuals who performed better in the difficult L2 rule were actually N400-dominant when processing word order violations in English. Beta desynchronization in the familiar L2 rule was also negatively correlated with RDI values ($r(50) = -0.31$, $p = 0.026$), showing that semantically driven individuals built stronger predictions for English word order. These results suggest what has already been argued by data from native speakers, that the P600 is not necessarily the marker of complete acquisition in all constructions or even in all languages. In fact, Chinese speakers showed N400 effects in response to word order violations in their L1 at the group level whereas English speakers did not. As a result, very different conclusions would be drawn if learners' ERP responses to the familiar rule were compared to native processing of the rule in Chinese versus in English at the group level. That is, whether the N400-response is considered native-like greatly depends on whether the native processing is done by Chinese or English speakers. Again, this is in addition to the fact that even some native speakers of English were found to be N400-dominant in the English word order condition.

Taken together with the previous literature, the current findings underscore a need to reconcile the relationship between increases in L2 proficiency being related to the development of the P600 effect with the status of the N400 as also being a native response to syntactic errors. One possibility is that all things being equal, the P600 may require more time to appear in the ERP waveform. Recall that early in acquisition, grammar knowledge is fairly idiomatic and tied to word forms. As a result, N400 responses are the norm for novice learners. Individuals who eventually become clearly P600-dominant may already be developing a P600 early on, but the overlap with the N400 component during averaging masks the presence of the effect. Consequently, P600 magnitude is typically correlated with proficiency and accuracy because it emerges out from under the shadow of the N400 over time as learners generally become more highly proficient over L2 exposure and use. In this way, meaning- and form-focused individuals may already have similar native-like performance before the P600 effect itself becomes apparent in the grand average waveform.

4.2 Variability in linguistic predictions as a feature of the comprehender

As interest in individual differences in language processing has increased in recent years, so too has the effort to understand where this variability comes from. Differences in language proficiency and working memory have been found to explain some of the variability observed during native language comprehension (Kim et al., 2018; Nakano et al., 2010; Pakulak & Neville, 2010; C. Weber-Fox et al., 2003). However, there is increasing evidence that suggests individual differences in comprehension may reflect a feature of the individual that is not fully explained by these cognitive or experience related factors (O’Rourke & Colflesh, 2015; Tanner, 2019; Tanner & Van Hell, 2014). Data from the current experiments add to this growing body of work: neither working memory capacity, language analytic ability, nor grammar proficiency explained variability in English or Chinese speakers’ electrophysiological responses or behavioral sensitivity to the syntactic or semantic errors in their native language. In fact, data from Experiment 1 suggest that the best predictor of variability in language processing is in fact variability in language processing: English speakers' syntax RDI values predicted semantic RDI values even when controlling for these explanatory factors, meaning that individual differences in linguistic prediction were related across semantic and syntactic anomalies over and above the contribution of cognitive and language experience variables. Although not of primary interest, additional analyses uncovered significant positive correlations between implausible RDI values and reading experience and vocabulary aptitude (measured by the LLAMA B subcomponent of the LLAMA Aptitude Test), as well as a marginal positive correlation with vocabulary proficiency in the English proficiency test (p's $< .058$). However, when these three variables were entered as predictors of implausible RDI values along with verb tense RDI, the significant model showed that only verb tense RDI was a significant predictor of individual differences in semantic anomalies ($F(3, 53) = 3.02$, $p = 0.038$, $R^2 = 0.10$). Thus, individuals appear to systematically deploy semantically or syntactically driven predictions across linguistic contexts and this characteristic of the comprehender does not seem to be influenced by individual differences in working memory, grammar or vocabulary aptitude or proficiency, or reading experience. It is therefore possible that had previous research included a measure of language processing variability along with working memory or language proficiency predictors, the language processing feature of the comprehender observed here may have also become evident.

4.2.1 *The influence of context on linguistic predictions*

To date, there have been few investigations into the degree to which individual differences in syntactic processing are consistent within an individual. Nevertheless, there appears to be some consensus in the literature thus far that this processing variability reflects a stable trait of the comprehender that is deployed across syntactic processing contexts (Bice &

Kroll, 2017; Finestrat et al., 2020; Tanner, 2019; Tanner & Van Hell, 2014). In these studies, the primary evidence for this conclusion is the finding of significant positive correlations between RDI values in the conditions of interest. Contrary to these results, however, there were no significant correlations between RDI values in either English or Chinese participants across syntactic conditions in the present studies. Nevertheless, variability in verb tense anomaly processing predicted variability in semantic anomaly processing. The current findings thus suggest that although there may be some systematicity to the way comprehenders respond to linguistic errors, the differences among speakers are not stable enough to be considered an invariable trait that manifests in all linguistic contexts.

Instead, it may be argued that these individual differences become evident in instances where linguistic predictions can be made with some certainty. When comprehenders become familiar with the structures in an experimental design and the context becomes more predictable, they may begin to deploy stronger predictions informed by semantic or syntactic cues made available before the critical word. There is supporting evidence that individuals' expectations can be affected even within a short experimental session. For instance, although comprehenders initially showed N400 effects to animacy violations such as "*The peanut was ?<u>in love</u>*" compared to "*The peanut was <u>salted</u>*", the N400 effect flipped after repeated exposure to contexts that supported an animate interpretation of *peanut* such that N400 magnitude became larger to "*The peanut was <u>salted</u>*" (Nieuwland & Van Berkum, 2006).

This may explain why RDI patterns were related across verb tense anomalies and semantic anomalies but not in the remaining syntactic conditions. In the syntax block, verb tense anomalies were presented simultaneously with word order and subject-verb agreement anomalies. Compared to the word order and collective noun conditions, the verb tense condition was the only violation to occur after a verb (e.g., "*Mandy was *<u>cook</u>…*" vs. "*Their niece ate soup *<u>her</u>*"/"*The committee *<u>have</u>…*"). Over the course of the experimental session, the auxiliary verb *was* may have begun to cue participants to pay close attention to the subsequent word to check whether it violated the expected agreement. As a result, comprehenders would have built stronger semantic or syntactic expectations for the critical word in this condition because it would have been advantageous to do so: given that the grammaticality judgment task required participants to determine whether a sentence was a well-formed sentence, the predictability of the context would have allowed participants to pre-activate expected continuations and make sentence comprehension and judgment more efficient. In contrast, the pre-critical words in the word order and collective noun conditions did not provide as strong of a cue about what might come next, so prediction would not have been as useful. In this way, both the local, sentence-level context and the global, experiment-level context, would have pushed participants to commit to more concrete predictions in the verb tense condition over the others. Indeed, the verb tense condition had far more variability in RDI values than the other who syntactic conditions, indicating that individuals relied more on semantic or syntactic cues in this context.

The semantic block was another context where prediction would have been advantageous for efficient processing. The sentence frames were constructed to be highly constraining in order to set up highly expected continuations. Exposure to these stimuli would have encouraged use of linguistic cues to predict upcoming input, even more so upon encountering unexpected or implausible continuations. Quickly integrating linguistic information and deploying strong predictions would have again facilitated comprehension as well as sentence judgment. Semantically driven participants would have thus been especially sensitive to the semantic context build up by the sentence frame, leading them to make stronger syntactic predictions that were violated to greater degrees in the unexpected and implausible conditions. As a result, semantically driven participants showed larger N400 magnitudes than syntactically driven individuals whose predictions for meaning were less strong.

For Chinese participants, the context that would have influenced comprehenders to build stronger predictions extended past the sentence or even experiment level to the broader context of residence in an English-speaking environment. The rules for word order in English are stricter than those in Chinese, so in order to complete the sentence judgment task successfully, it would be beneficial to attend more closely to the syntactic structure and build form-based predictions in both the word order and the verb tense conditions. Even in the verb tense condition, awareness of word order would permit learners to make use of the auxiliary verb *was* to predict the appropriate upcoming agreement morpheme *-ing* (e.g., "*Mandy was <u>cooking</u>…*"). Those learners who build stronger syntactic predictions would thus be more efficient at detecting a verb tense agreement error in order to make the correct behavioral response. As the data show, P600-dominant Chinese speakers did in fact perform better on the sentence judgment task in the English verb tense condition compared to N400-dominant Chinese speakers whose syntactic predictions were not as strong.

The influence of context on the elicitation of language processing variability may also explain instances where RDI values were found to be positively correlated in the previously cited literature (Bice & Kroll, 2017; Finestrat et al., 2020; Tanner, 2019; Tanner & Van Hell, 2014). First, in all of these studies, RDI values were compared across stimuli that violated rules of morphosyntactic agreement. Tanner and van Hell (2014) compared verb tense errors, such as the ones found in the current experiments (e.g., *The crime rate was *<u>increase</u>*), to subject-verb agreement errors (e.g., *The roses *<u>grows</u> in the garden*) and found significant positive correlations. Tanner (2019) found that RDI values correlated across subject-verb agreement errors that contained either lexical verbs (e.g. *grows*) or auxiliary verbs (e.g., *was*). Bice and Kroll (2017) showed correlations in subject-verb agreement errors across L1 English and L2 Spanish, and Finestrat et al. (2020) additionally showed correlations in noun phrase agreement errors. In all four cases, linguistic constructions supported the need to build predictions for appropriate agreement, a pattern that could have become fairly evident given that comprehenders were exposed to only one or at most two types of syntactic violations within one session, as is common in most ERP experiments. For example, analysis of stimuli from Tanner (2019) reveal that all sentences in the critical condition started with *The NOUN* followed by a verb and the other filler syntactic condition (which violated reflexive gender agreement, e.g., *The anxious cowboy prepared <u>himself</u>/*<u>herself</u>…*) always started with *The ADJ NOUN* (Tanner, 2018). Within the second word of the sentence, a comprehender could determine what condition the sentence was in and make a prediction about the appropriate verb forms that should follow. Therefore, including several different syntactic violation types in one session may have made it generally more difficult for English speakers to predict the condition a sentence was in in Experiment 1, except for the case when an auxiliary verb signaled the need to predict for the purpose of grammatical agreement. In fact, the magnitude of the N400 effect in the verb tense condition was more prominent than the one observed in Tanner and van Hell (2014), where participants were exposed to only two syntactic conditions. Since the verb tense condition was more varied from the other two syntactic conditions, participants may have deployed stronger predictions in the current experiment compared to participants who processed verb tense errors in Tanner and van Hell (2014).

4.2.2 *Variability in linguistic predictions: state vs. trait*

It has been argued thus far that variability in linguistic predictions does not appear to be a stable trait that occurs across all linguistic contexts and may instead manifest primarily in highly predictable contexts where prediction becomes particularly advantageous for the comprehender. This reasoning supports a state interpretation for the linguistic prediction feature, that the reliance on semantic or syntactic expectations for comprehension takes place under certain conditions. It is as yet unclear how participants arrive toward a preference for one type of prediction over another. One possibility is that comprehenders make the decision dynamically within the context of the experiment. Perhaps a choice is made early in the experiment that proves to be effective in making predictions and feedback strengthens the choice into a strategy. This strategy is subsequently retained in the second experimental session and leads to the observed similarity across conditions. As a result, participants who find semantic or syntactic predictions useful continue to rely on them for comprehension, whereas participants whose RDI values hover around 0, indicating equally larger (or equally absent) N400 and P600 ERP effects, do not choose one strategy over the other. Therefore, some combination of experiment and participant characteristics may lead to preference for one type of linguistic prediction over another.

Another possibility is that comprehenders already arrive with some predisposition to rely on certain linguistic expectations that the context brings to light. Predictions cannot be made without some database of information on which to build those predictions. When predictions are supported, the importance of that information in the database is strengthened. Some contexts may thus increase the relevance of particular linguistic cues and individuals who attend more to those cues therefore deploy more accurate predictions. This may in fact have been the case for

Chinese participants in Experiment 2, where the broader English-speaking environment would have strengthened the importance of syntactic cues for the purpose of understanding English word order and verb tense agreement. As already discussed, P600-dominant Chinese speakers who built stronger syntactic predictions in Chinese performed better on the sentence judgment task in the complex English verb tense rule.

However, the influence of the English-speaking environment on linguistic prediction may actually extend beyond the context of English language comprehension. The discussion thus far has assumed that the relationship between L1 processing and L2 learning is generally unidirectional, such that features of a learners' L1 drive differences in L2 acquisition. However, given that similar cognitive and neurological mechanisms underpin L1 and L2 systems, there is no reason why the relationship could not in fact be bidirectional. There is indeed evidence of L2 to L1 transfer at various levels of linguistic analysis (Bergmann et al., 2016; Brown & Gullberg, 2008; Bylund & Jarvis, 2011; Hohenstein et al., 2006; Hwang et al., 2018; Muñoz & Cadierno, 2019; Ulbrich & Ordin, 2014). In fact, findings from the current data suggest that Chinese speakers' experience with English may have affected their L1 processing patterns. First, Chinese word order RDI was significantly correlated not just with English verb tense a-prime but also with Chinese word order a-prime ($r(50) = 0.39$, $p = 0.006$). That is, Chinese speakers who were P600-dominant in response to errors in their native language were also more sensitive to those errors. Compare this result to data from English speakers, where sentence judgement performance was not related to processing variability in any syntactic or semantic condition.

Additional analyses not of primary interest also revealed that variability in Chinese word order beta desynchronization was positively correlated with total years of English experience ($r(50) = 0.34$, $p = 0.016$) as well as English grammar proficiency ($r(50) = 0.33$, $p = 0.018$).

Thus, the more Chinese speakers gained experience and proficiency in English, the more syntactic errors interrupted their top-down native language predictions, suggesting that learners' expectations of Chinese word order were strengthened with exposure to the L2. It has been argued that word order in Chinese is governed by informational and communicative constraints, meaning that the order of words and phrases is determined based on the message that is to be imparted rather than on the grammatical functions that words play. This is in contrast to English, where word order is determined quite strictly by the grammatical functions of words. Thus, word order is understood to be more strict in English than it is in Chinese. As a result, it is possible that Chinese speakers who had more experience with English and its strict word order became more sensitive to word order violations in their native language. Given Chinese participants' daily use of English and residency in an English-speaking environment during the time of testing, the state these comprehenders were in would have supported the focus on syntactic information and resulted in better performance in those individuals who attended to available syntactic cues to build their predictions.

It is of course possible to argue that a predisposition to build semantic or syntactic expectations is simply a trait of the comprehender determined prior to exposure to the context of the experimental task, as previous research has posited. There is perhaps some evidence in support of the existence of a more stable trait when variability in P600 effect magnitude is used as a dependent variable instead of the RDI. When P600 magnitudes were correlated across conditions, there was a significant positive correlation between P600 effect magnitude in the word order and verb tense conditions ($r(55) = 0.30$, $p = 0.023$) and a marginal positive correlation between the word order and collective noun conditions ($r(55) = 0.25$, $p = 0.066$) in Experiment 1. There was also a significant positive correlation between P600 effect magnitude in

English word order and English verb tense in Experiment 2 ($r(50) = 0.43$, $p = 0.002$), as well as a significant positive correlation between Chinese word order and English word order when all trials are included in the computation of the P600 ($r(50) = 0.29$, $p = 0.041$). It is possible, then, that degree of sensitivity to violations of grammatical expectations, as measured by the P600, is more context-agnostic than the degree of overall strength of semantic or syntactic predictions, as measured by the RDI. The P600 effect is a direct measure of neural response to syntactic anomalies, whereas the RDI is a derived measure of competition between the processes that elicit N400 and P600 effect magnitudes. This might explain why the extant literature has consistently reported that syntactic errors elicit P600 responses at the group level. Many, if not all, comprehenders would typically show a P600 response, but the P600 effect may not necessarily be the dominant response for all comprehenders. The RDI may thus be best suited for tracking individual differences in contexts where reliance on semantic or syntactic prediction becomes advantageous and in individuals who choose one strategy over the other. Given that all linguistic input has some level of semantic and syntactic information, it is likely that comprehenders are always deploying some degree of semantic and syntactic predictions. When the N400 and P600 magnitudes are analyzed alone, they may reflect some baseline strength of semantic or syntactic expectations within an individual. However, when the context encourages stronger prediction, one type of expectation may become more useful for the comprehender, and this preference is what is captured by the RDI.

4.3 The influence of the L1 processing variability on L2 acquisition

In general, the ERP evidence from Experiment 2 aligns with previous research on the effect of L1-L2 similarity on L2 acquisition (Chen et al., 2007; Kotz, 2009; Ojima et al., 2005; Sabourin & Stowe, 2008). Chinese speakers demonstrated better behavioral performance and more native-like ERP patterns to the familiar L2 rule compared to the unique L2 rule, even after an average of 10 years of exposure to English in an instructional setting or as residents of the U.S. However, not all participants struggled with the unique L2 rule to the same degree. Adult L2 acquisition is a highly complex task and numerous learner characteristics have been implicated in L2 learning success, including motivation, language aptitude, age of acquisition, and learning style, among many others (Dörnyei & Ryan, 2015; Dörnyei & Skehan, 2003; Skehan, 1998; Tanner et al., 2014). Data from the current experiment provide support for the role of a highly understudied variable, the learners' L1 linguistic prediction feature. Chinese learners' increased behavioral sensitivity to a complex, unique L2 rule was predicted by more positive RDI values in response to syntactic violations in L1, over and above factors related to L2 experience or aptitude. Thus, by building stronger syntactic predictions, P600-dominant individuals may be better at acquiring new syntactic forms in an L2.

These findings are in line with evidence from research on beta desynchronization and error-related negativities that better syntactic predictors make better syntactic learners (Bastarrika & Davidson, 2017; Batterink & Neville, 2014; Bultena et al., 2017; Cross et al., 2020; Davidson & Indefrey, 2007, 2009, 2011; Misyak et al., 2010). If an individual builds stronger expectations of form in the L1 and has a database that weights syntactic information to a greater degree, they may be more likely to attend to syntactic cues in L2 compared to semantically driven individuals. By attending early on to syntactic cues, grammatically focused learners begin to build form-based predictions more quickly. When input conflicts with those expectations, prediction errors are deployed to adapt those expectations, thereby refining form-based predictions in future communication. In this way, P600-dominant individuals may excel at acquiring L2 unique rules compared to their N400-dominant counterparts. However, N400-

dominant individuals who focus on semantic cues and make meaning-based predictions may show greater success in learning vocabulary from context, although this remains to be tested. Nevertheless, the distinction observed between semantically and syntactically driven comprehenders may help explain why some students seem to particularly excel in classrooms that stress form-focused, explicit instruction of grammatical rules, while others prefer environments that emphasize meaning-focused, implicit instruction (e.g. immersion) (Skehan, 1998). By assessing individual differences in L1 language processing, it would therefore be possible to ascertain whether an L1 learner will succeed in form- versus meaning-based pedagogy prior to instruction and thus leverage this feature of the comprehender to increase individual L2 learning outcomes.

However, there are two limitations to the interpretation that L1 processing variability influences L2 acquisition. The first, as already discussed, is the possibility that it was the acquisition of the L2 that influenced the L1 linguistic prediction feature in Chinese participants, although the primarily supporting evidence stems from follow-up analyses of correlations with beta desynchronization, not from ERP data. The second limitation is that the effect of Chinese speakers' L1 language processing variability predicted differences in a-prime scores in the unique L2 rule, not in RDI values. Just as in Experiment 1, there were no significant correlations between RDI values in any of the conditions Chinese speakers were exposed to. Although this is contrary to the few available findings comparing RDI across languages (Bice & Kroll, 2017; Finestrat et al., 2020), there was in fact a positive correlation between P600 magnitudes in Chinese word order and English word order when all trials were included in the analysis ($r(50) = 0.29$, $p = 0.041$). Even if the environmental and experimental contexts encouraged participants to attend more to word class information as a cue for word order expectations, it is still fairly difficult to predict a specific class of words that might follow a noun (e.g., *Their niece ate soup…*). As a result, there would be less competition between semantic and syntactic prediction processes, as both would have weak predictions at best, and thus no relationship between RDI would be detected. On the other hand, it would be easy to detect a word order anomaly as being a violation of grammatical expectations, as measured by the P600: while it might be difficult to predict with high certainty what type of word could follow a noun (e.g., preposition, adverb, conjunction, determiner, etc.: *Their niece ate soup outside/quickly/and…/every (day)…*), it is much easier to predict with high certainty what word class cannot follow a noun (e.g., pronoun: *Their niece ate soup *<u>her</u>…*). Consequently, variability in P600 magnitudes were related in word order violations across languages, but RDI values were not. Finally, with regard to the unique L2 rule, recall that the correlation between N400 and P600 effects in the English verb tense condition was positive, indicating that Chinese learners were still developing dominance patterns for this condition. RDI values in the English verb tense condition therefore hovered around 0 for the majority of participants, making it impossible to detect a significant correlation in RDI values with either Chinese word order or English word order, where RDI values ranged from strongly N400-dominant to strongly P600-dominant (mean RDI range = -8.9 to 10.1) (see Figure 18 and Figure 19).

4.4 Individual differences in beta desynchronization

In all conditions across both experiments, syntactic and semantic violations resulted in lower beta power, as has previously been reported (Bastiaansen et al., 2010; Kielar et al., 2014; Lewis & Bastiaansen, 2015). Although there was variability in the degree of beta desynchronization across individuals, there were few observed relationships between neural oscillations and ERP effects. Nevertheless, the existing correlations do provide additional evidence in support of the prediction account of individual differences in language processing. Mirroring the inverse relationship found between syntactic RDI and semantic RDI values, degree of beta desynchronization in the word order condition was revealed to be negatively correlated with implausible beta desynchronization in English participants. That is, the more a syntactic violation interrupted a comprehender's maintenance of the top-down sentence level representation, the less that comprehender's top-down representation was affected by a semantic anomaly. On the other hand, individuals whose top-down contextual predictions were less affected by syntactic anomalies appeared instead to interrupt maintenance of their top-down representations to a greater degree when a semantic continuation did not meet their expectations. This is consistent with the idea that semantically focused individuals would be more likely to predict the highly expected continuations supported by the strongly constraining sentence contexts, and thus an implausible continuation would lead to a robust decrease in beta, reflecting the dismantling of the established cognitive net supporting the strong top-down expectation.

In addition, word order beta desynchronization was also found to be positively correlated with word order RDI values: the more syntactic errors violated grammatical predictions in an individual, the more that individual interrupted maintenance of their top-down contextual predictions for the sentence under consideration. This is in line with previous research showing that P600 magnitude and beta desynchronization are positively correlated (Schneider & Maguire, 2018). However, it is not clear that P600 magnitudes are always related to larger degrees of beta desynchronization. In Chinese speakers, there was a negative correlation between RDI values and beta desynchronization in English word order ($r(50) = -0.31$, $p = 0.026$). In this condition, it was N400-dominant individuals that built stronger top-down predictions, indicating that English word order errors resulted in greater violations of expectation in semantically driven individuals.

Given that different language contexts showed different relationships between RDI values and beta desynchronization, this can be taken as further evidence that variability in language processing may very well be context dependent. Nevertheless, consistent with individual differences analyses of beta overall, beta desynchronization may not differ across participants to the same degree as ERPs do, lending support to the finding that comprehenders generally arrive at equally strong predictions regardless of whether those expectations are semantically or syntactically driven.

4.5 Implications

4.5.1 *Individual differences in the communication model of comprehension*

Taken together, the findings from the present experiments reveal that individuals vary with regard to the strength of their semantic or syntactic predictions during language comprehension. Given that the data support a prediction-based explanation for individual differences in language processing, these findings can further be interpreted in light of the communication model of comprehension proposed by Kuperberg and colleagues (2020). According to this proposal, the N400 indexes the degree to which incoming words match predicted features at the semantic feature level, the lowest level of the communication model hierarchy. The P600, on the other hand, indexes the degree to which events and actions at the intermediate event level can occur under the predictions of the situation model, the highest level of the communication model hierarchy. The individual differences data suggests then that individuals who attend to semantic information for prediction build more semantically refined expectations and pre-activate more semantic features in anticipation of the incoming input. When a semantically anomalous word appears, its features get checked against the preactivated features at the semantic level and the large degree of mismatch results in a large N400 effect. A comprehender who builds form-based expectations, on the other hand, pre-actives fewer semantic features, resulting in less of a mismatch at the semantic level and thus a smaller N400 effect. For instance, in the sentence "*Finally at home, Perry took out his keys to unlock the…*", a semantically driven individual may already pre-activate all features of the word *door* (e.g., <unlockable>, <entrance>, <residential>, <inanimate>, <stationary>, etc.) whereas a syntactically driven individual may pre-active broad features such as <unlockable> that coincide with intermediate level event structure expectations that Perry will unlock something. Upon encountering an unexpected word (e.g., *kennel*) or an implausible word (e.g. *traffic*), the syntactically driven comprehender will have less of a mismatch of their semantic expectations than the semantically driven comprehender because they activate fewer features at the semantic level.

Under the communication model of comprehension, linguistic predictions in syntactically driven individuals manifest at the intermediate level of the event model rather than at the lowest level of the semantic feature model in semantically driven individuals. For these individuals, unexpected and implausible words pass through the semantic level without much need for adjusting pre-activated features. It is at the event and situation levels, however, where mismatches with predictions are addressed for these comprehenders. By attending more to syntactic cues, these individuals instead focus their predictions on the events and actions that can take place under the situation model. They check to ensure that the incoming input matches expectations regarding an event structure where the agent *Perry* unlocks some patient. Thus, when the parser encounters *traffic*, a non-viable patient for this event, the P600 reflects a larger degree of mismatch between the event that was expected to take place and the event under consideration. For semantically driven individuals whose predictions manifest at the semantic level, the expectations at the event level are not as strictly defined, and they thus elicit larger smaller P600 effects.

For comprehenders who build syntactic predictions, unexpected continuations must also be reckoned with at the event and situation levels. In fact, syntactically driven individuals not only showed a larger later frontal positivity than people who built semantic predictions, but they also showed a P600 effect where semantically driven individuals did not. This indicates that individuals who build syntactically informed predictions show a greater degree of mismatch between their situation model and the event model, as reflected by the P600, as well as a greater amount of updating of the situation model, as reflect by the frontal positivity. It is possible that if these individuals do not attend to semantic information to the same degree as their semantically driven counterparts, their high-level situation model is not as informed or detailed. Upon encountering an unexpected continuation like *kennel*, the situation model requires more updating to flesh out the comprehender's interpretation of the sentence meaning under construction. The mismatch between the situation model and its possible events may thus also be greater if the situation model is underspecified and does not yet accommodate events that could involve *kennel*.

If the finding that P600-dominant and N400-dominant individuals differ in the type of linguistic predictions they build is interpreted in light of the communication model, then it would appear that individuals deploy their linguistic predictions at different levels of the communication hierarchy. This may explain the tradeoff observed between N400 and P600 effects in the current experiments as well as in previous literature, that N400 and P600 effect magnitudes are found to be negatively correlated with one another (Grey et al., 2017; Kim et al., 2018; Meulman et al., 2014; Qi et al., 2017; Tanner, 2019; Tanner et al., 2014; Tanner & Van Hell, 2014). The inverse relationship between these effects may thus reflect the degree to which an individual recruits different levels of the communication hierarchy for representing and monitoring their predictions. If an individual primarily deploys structural predictions, for instance, an anomaly will mainly be addressed at the event level and will inform the situation model without much interaction with the semantic level, resulting in a P600 effect but not an N400 effect. Comprehenders at the lower extreme of the RDI spectrum may therefore strongly prefer the semantic level over the event level, whereas comprehenders with high RDI values may prefer the reverse.

4.5.2 *Implications for research on L1 comprehension*

By exposing comprehenders to a number of diverse linguistic anomalies, the current experiment was better able to characterize individual differences in language comprehension from a perspective not previously assessed by research in this domain. The findings underscore that variability in language processing is the rule, not the exception. If ERP research is to inform models of language production and comprehension, it should not continue to do so without investigating the variability that underlies the mean. Furthermore, not only do comprehenders behave in considerably different ways within a single linguistic construction, but that behavior can further change across linguistic contexts, sometimes in unpredictable ways. As a result, stimuli should be selected with caution as the type of linguistic anomaly chosen could affect how another linguistic structure is processed and also severely limit generalizability of findings. For instance, although verbal working memory has been shown to predict individual differences in processing semantic attraction anomalies such as "*The hearty meal was ?<u>devouring</u>*," or animacy violations (Kim et al., 2018; Nakano et al., 2010), it has not been shown to reliably predict variability in violations such as the ones tested in the current experiment (O’Rourke & Colflesh, 2015; Tanner, 2019). This underscores the need for presenting participants with varied stimuli if the goal of the research is to understand individual differences within and across languages.

In addition, ERP researchers should consider taking more advantage of the versatility of the EEG signal and perform time-frequency analyses in combination with the computation of ERP components. Time-frequency analyses provide insight into network-level behavior related to the synchronization and desynchronization of neural resources involved in carrying out various cognitive processes, without the need for any additional data collection efforts. As a result, ERP effects and time-frequency analyses can inform the study of language comprehension from quite different perspectives. For example, in the current experiment, unexpected and expected semantic continuations did not differ in their degree of beta desynchronization even though unexpected continuations showed larger N400 magnitudes than expected continuations. This suggests that although unexpected stimuli violated semantic predictions, they were still accepted as being consistent with the ongoing top-down representation of the sentence meaning. By examining the same stimuli via complementary analyses, the data reveal nuances in language processing with regard to how different linguistic mechanisms contribute to resolving the challenge of comprehension.

4.5.3 *Implications for research on L2 acquisition*

Perhaps the strongest takeaway for research on individual differences in L2 learning is that the P600 effect is not the only native-like ERP response to syntactic violations. In every syntactic condition across English and Chinese stimuli, native speakers showed N400-dominant responses to grammatical violations, indicating that N400 responses are just as native-like as P600 responses. As a result, the absence of a P600 effect should not be taken as proof positive of non-native grammatical processing. Although the RDI is useful for tracking the variability found in native speakers' ERP responses, its use as a measure of native-like processing is still limited for two reasons. First, given that native responses fall along the entire RDI continuum, it is difficult to assess whether a learner is in the process of achieving native-like competency or has already achieved native-like competency based on the learners position on the RDI spectrum without tracking that individual longitudinally. Second, although there were no differences in RDI values across Chinese or English conditions in Experiment 2, the underlying relationship between N400 and P600 magnitudes was quite different between the familiar L2 rule and the unique L2 rule. Chinese speakers showed a strong negative correlation in English word order, just as native English speakers did, but the relationship between these effects was positive in English verb tense. This was in contrast to the negative correlation found in native English speakers for this rule. Together with the lower a-prime scores in that condition, this opposite pattern indicated that Chinese learners had not yet acquired native-like proficiency in the unique L2 rule.

Consequently, instead of using grand average ERP waveforms or even RDI values, future research on L2 acquisition should use the correlation between native speakers' N400 and P600 magnitudes as a basis of comparison to determine whether L2 learners are capable of acquiring native-like proficiency in a grammatical rule. Whenever this correlation has been reported in native speakers, there has been a strong negative relationship between ERP effects (Finestrat et al., 2020; Grey et al., 2017; O'Rourke & Colflesh, 2015; Tanner, 2019; Tanner et al., 2014). Given this consistent finding, positive or zero correlations between ERP effects may indicate that a group of learners is still developing native-like sensitivity to an L2 rule. The strategy for assessing the degree to which an individual is approaching native-like performance is less clear, although it is likely to require taking a-prime sensitivity into account in combination with ERP responses. Additional longitudinal research characterizing how the N400 vs. P600 correlation changes over acquisition would provide insight into how an individual moves across the continuum and how learner characteristics differentially influence this transition.

Time-frequency analyses could also provide important insight for studies of language acquisition. In the present study, Chinese participants showed similar degrees of beta desynchronization across all Chinese and English conditions even though the ERP data were characterized by considerably different patterns. By only examining ERP responses, the data leads to the conclusion that Chinese speakers have not acquired native-like sensitivity to the L2 unique rule even after an average of 10 years of exposure to and experience with English. However, given that violations of the familiar L2 rule and the unique L2 rule interrupted maintenance of top-down predictions to the same degree, and furthermore did not differ from responses to violations in the learners’ L1, this interpretation should be taken with some caution. It appears that Chinese learners were in fact sensitive to the unique L2 rule but they may have struggled to deploy specific semantic or syntactic predictions quickly enough to show violations of expectations in the ERP effects even though they were consciously aware of the error during online processing. This is in line with the literature on L2 grammatical gender acquisition, which poses great difficulty for native English speakers to acquire. Hopp (2013) found that L2 German speakers failed to show native-like ERP responses to grammatical gender violations even after having lived in Germany and used German on a daily basis for 30 years. Despite the absence of ERP effects, learners showed high accuracy when labeling the gender of a noun and identifying incorrect gender agreement in behavioral responses. Hopp (2013) thus proposed that due to the absence of this unique L2 rule in the L1, learners struggled to extract the gender of the noun from its lexical entry and quickly map it to the surface structure of the sentence during online language comprehension. Similarly, Chinese speakers in this experiment were in fact sensitive to the L2 unique rule, but this finding would not be evident had beta desynchronization not been assessed in addition to the ERP effect magnitudes. Any investigators who have previously drawn conclusions about language acquisition based on ERP findings could return to the very same EEG recordings and re-analyze the data using time-frequency analyses to further inform their research.

4.6 Conclusion

In summation, the experiments reported herein provide evidence that individual differences in language processing are borne out of variability with regard to whether comprehenders built semantic or syntactic predictions during comprehension. By assessing a variety of linguistic anomalies in L1 and L2 using two different analyses of the EEG signal and controlling for a number of cognitive and language experience variables, the data support the existence of a language processing feature that interacts with the sentence level, experimental, and even environmental contexts to drive individuals to dynamically address the challenge of language comprehension from different directions. Moreover, the data consistently point to the fact that variability in language processing is the rule, not the exception. This has considerable implications for understanding language processing across the entire research domain: individual differences exist in children, patients, deaf speakers, and even typical, healthy adults. With greater attention to variation in linguistic predictions, a growing understanding of these individual differences can be leveraged to help increase language comprehension, language learning, and ultimately lifelong success in all people.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The Robustness of Aptitude Effects in Near-Native Second Langauge Acquisition. *Studies in Second Language Acquisition*, *30*(4), 481–509. JSTOR.

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica*, *128*(3), 466–478. https://doi.org/10.1016/j.actpsy.2008.03.014

Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods*, *40*(1), 278–289. https://doi.org/10.3758/BRM.40.1.278

Alboukadel, K. (2020). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests* (R package version 0.4.0) [Computer software]. https://CRAN.R-project.org/package=rstatix

Andringa, S. (2014). The use of native speaker norms in critical period hypothesis research. *Studies in Second Language Acquisition*, *36*(3), 565–596. JSTOR. https://doi.org/10.2307/26330708

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Arnal, L. H., Wyart, V., & Giraud, A.-L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience*, *14*(6), 797–801. https://doi.org/10.1038/nn.2810

Bastarrika, A., & Davidson, D. J. (2017). An Event Related Field Study of Rapid Grammatical Plasticity in Adult Second-Language Learners. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00012

Bastiaansen, M., & Hagoort, P. (2006). Oscillatory neuronal dynamics during language comprehension. In C. Neuper & W. Klimesch (Eds.), *Progress in Brain Research* (Vol. 159, pp. 179–196). Elsevier. https://doi.org/10.1016/S0079-6123(06)59012-0

Bastiaansen, M., & Hagoort, P. (2015). Frequency-based Segregation of Syntactic and Semantic Unification during Online Sentence Level Language Comprehension. *Journal of Cognitive Neuroscience*, *27*(11), 2095–2107. https://doi.org/10.1162/jocn_a_00829

Bastiaansen, M., Magyari, L., & Hagoort, P. (2010). Syntactic Unification Operations Are Reflected in Oscillatory Dynamics during On-line Sentence Comprehension. *Journal of Cognitive Neuroscience*, *22*(7), 1333–1347. https://doi.org/10.1162/jocn.2009.21283

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical Microcircuits for Predictive Coding. *Neuron*, *76*(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038

Batterink, L., & Neville, H. J. (2014). ERPs Recorded during Early Second Language Exposure Predict Syntactic Learning. *Journal of Cognitive Neuroscience*, *26*(9), 2005–2020. https://doi.org/10.1162/jocn_a_00618

Bergmann, C., Nota, A., Sprenger, S. A., & Schmid, M. S. (2016). L2 immersion causes non-native-like L1 pronunciation in German attriters. *Journal of Phonetics*, *58*, 71–86. https://doi.org/10.1016/j.wocn.2016.07.001

Berkum, J. J. A. van, Hagoort, P., & Brown, C. M. (1999). Semantic Integration in Sentences and Discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*(6), 657–671. https://doi.org/10.1162/089892999563724

Bice, K., & Kroll, J. F. (2017). *Using ERPs to investigate the stability of individual differences in sentence processing in two languages*. Bilingualism Matters Inaugural Conference, University of California-Riverside.

Birdsong, D. (2006). Age and Second Language Acquisition and Processing: A Selective Overview. *Language Learning*, *56*(s1), 9–49. https://doi.org/10.1111/j.1467-9922.2006.00353.x

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. M. Gass & J. Schachter (Eds.), *Linguistic Perspectives on Second Language Acquisition* (1st ed., pp. 41–68). Cambridge University Press. https://doi.org/10.1017/CBO9781139524544.005

Bond, K., Gabriele, A., Fiorentino, R., & Bañón, J. A. (2011). Individual Differences and the Role of the L1 in L2 Processing: An ERP Investigation. In J. Herschensohn & D. Tanner (Eds.), *Proceedings of the 11th Generative Approaches to Second Language Acquisition Conference (GASLA 2011)* (pp. 17–29). Cascadilla Proceedings Project.

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is Enough: N400 Indexes Semantic Integration of Novel Word Meanings from a Single Exposure in Context. *Language Learning and Development : The Official Journal of the Society for Language Development*, *8*(3), 278–302. https://doi.org/10.1080/15475441.2011.614893

Bressler, S. L., & Richter, C. G. (2015). Interareal oscillatory synchronization in top-down neocortical processing. *Current Opinion in Neurobiology*, *31*, 62–66. https://doi.org/10.1016/j.conb.2014.08.010

Brown, A., & Gullberg, M. (2008). Bidirectional Crosslinguistic Influence in L1-L2 Encoding of Manner in Speech and Gesture: A Study of Japanese Speakers of English. *Studies in Second Language Acquisition*, *30*(2), 225–251. JSTOR.

Bultena, S., Danielmeier, C., Bekkering, H., & Lemhöfer, K. (2017). Electrophysiological Correlates of Error Monitoring and Feedback Processing in Second Language Learning.

Frontiers in Human Neuroscience, *11*. https://doi.org/10.3389/fnhum.2017.00029

Bylund, E., & Jarvis, S. (2011). L2 effects on L1 event conceptualization. *Bilingualism: Language and Cognition*, *14*(1), 47–59. https://doi.org/10.1017/S1366728910000180

Canette, L.-H., Fiveash, A., Krzonowski, J., Corneyllie, A., Lalitte, P., Thompson, D., Trainor, L., Bedoin, N., & Tillmann, B. (2020). Regular rhythmic primes boost P600 in grammatical error processing in dyslexic adults and matched controls. *Neuropsychologia*, *138*, 107324. https://doi.org/10.1016/j.neuropsychologia.2019.107324

Carreiras, M., Salillas, E., & Barber, H. (2004). Event-related potentials elicited during parsing of ambiguous relative clauses in Spanish. *Cognitive Brain Research*, *20*(1), 98–105. https://doi.org/10.1016/j.cogbrainres.2004.01.009

Carroll, J. B. (1965). The Contributions of Psychological Theory and Educational Research to the Teaching of Foreign Languages. *The Modern Language Journal*, *49*(5), 273–281. JSTOR. https://doi.org/10.2307/322133

Carroll, J. B., & Sapon, S. M. (1959). *Modern Language Aptitude Test*. Second Language Testing Incorporated.

Chen, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition*, *10*(2), 161–174. https://doi.org/10.1017/S136672890700291X

Chow, W.-Y., & Phillips, C. (2013). No semantic illusions in the “Semantic P600” phenomenon: ERP evidence from Mandarin Chinese. *Brain Research*, *1506*, 76–93. https://doi.org/10.1016/j.brainres.2013.02.016

Clahsen, H., Lück, M., & Hahne, A. (2007). How children process over-regularizations: Evidence from event-related brain potentials. *Journal of Child Language*, *34*(3), 601–

622. https://doi.org/10.1017/S0305000907008082

Cross, Z. R., Corcoran, A. W., Schlesewsky, M., Kohler, Mark. J., & Bornkessel-Schlesewsky, I. (2020). *Oscillatory and aperiodic neural activity jointly predict grammar learning* [Preprint]. Neuroscience. https://doi.org/10.1101/2020.03.10.984971

Dąbrowska, E. (2018). Experience, aptitude and individual differences in native language ultimate attainment. *Cognition*, *178*, 222–235. https://doi.org/10.1016/j.cognition.2018.05.018

Dąbrowska, E. (2019). Experience, Aptitude, and Individual Differences in Linguistic Attainment: A Comparison of Native and Nonnative Speakers. *Language Learning, 69*(S1), 72–100. https://doi.org/10.1111/lang.12323

Davidson, D. J., & Indefrey, P. (2007). An inverse relation between event-related and time–frequency violation responses in sentence processing. *Brain Research, 1158,* 81–92. https://doi.org/10.1016/j.brainres.2007.04.082

Davidson, D. J., & Indefrey, P. (2009). An Event-related Potential Study on Changes of Violation and Error Responses during Morphosyntactic Learning. *Journal of Cognitive Neuroscience, 21*(3), 433–446. https://doi.org/10.1162/jocn.2008.21031

Davidson, D. J., & Indefrey, P. (2011). Error-Related Activity and Correlates of Grammatical Plasticity. *Frontiers in Psychology, 2*. https://doi.org/10.3389/fpsyg.2011.00219

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 600 million words, 1990-present.* www.english-corpora.org/coca/

DeKeyser, R. M. (2000). The Robustness of Critical Period Effects in Second Langauge Acquisition. *Studies in Second Language Acquisition, 22*(4), 499–533. JSTOR.

DeKeyser, R. M. (2008). Implicit and Explicit Learning. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313–348). Blackwell Pub.

DeKeyser, R. M., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, *31*(3), 413–438. https://doi.org/10.1017/S0142716410000056

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, *61*, 150–162. https://doi.org/10.1016/j.neuropsychologia.2014.06.016

Delong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, *48*(9), 1203–1207. https://doi.org/10.1111/j.1469-8986.2011.01199.x

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. https://doi.org/10.1038/nn1504

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dörnyei, Z., & Ryan, S. (2015). Language Aptitude. In *The psychology of the language learner revisited* (pp. 35–71). Routledge.

Dörnyei, Z., & Skehan, P. (2003). Individual Differences in Second Language Learning. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589–630). Blackwell Pub.

Douglas, R. J., & Martin, K. A. C. (1991). A functional microcircuit for cat visual cortex. *The Journal of Physiology*, *440*(1), 735–769. https://doi.org/10.1113/jphysiol.1991.sp018733

Douglas, R. J., Martin, K. A. C., & Whitteridge, D. (1989). A Canonical Microcircuit for Neocortex. *Neural Computation*, *1*(4), 480–488. https://doi.org/10.1162/neco.1989.1.4.480

Draheim, C., Harrison, T. L., Embretson, S. E., & Engle, R. W. (2018). What item response theory can tell us about the complex span tasks. *Psychological Assessment*, *30*(1), 116–129. https://doi.org/10.1037/pas0000444

Early Skills and Predictors of Academic Success. (2016). Hanover Research.

Engel, A. K., & Fries, P. (2010). Beta-band oscillations—Signalling the status quo? *Current Opinion in Neurobiology*, *20*(2), 156–165. https://doi.org/10.1016/j.conb.2010.02.015

English Proficiency Test. (n.d.). https://www.transparent.com/learn-english/proficiency-test.html

Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*, *41*(4), 469–495. https://doi.org/10.1006/jmla.1999.2660

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84. https://doi.org/10.1016/j.brainres.2006.06.101

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203. https://doi.org/10.1016/S0010-0285(03)00005-7

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Ferreira, F., & Clifton, C. (1986). The Independence of Syntactic Processing. *Journal of Memory and Language; New York*, *25*(3), 348–368.

Ferreira, F., & Patson, N. D. (2007). The ‘Good Enough’ Approach to Language Comprehension. *Language and Linguistics Compass, 1*(1–2), 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x

Finestrat, I., Abugaber, D., Luque, A., & Morgan-Short, K. (2020). *Do first language neural processes for morphosyntax transfer to the second language? Using event-related potentials to expand the evidence of cross-linguistic influence*. 1.

Fisher, K. J., Bassok, M., & Osterhout, L. (2010). When Two plus Two Does Not Equal Four: Event-Related Potential Responses to Semantically Incongruous Arithmetic Word Problems. *Proceedings of the Annual Meeting of Cognitive Science Society,* 1571–1576.

Fontolan, L., Morillon, B., Liegeois-Chauvel, C., & Giraud, A.-L. (2014). The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nature Communications, 5*(1), 4694. https://doi.org/10.1038/ncomms5694

Frazier, L., & Clifton, C. (1997). Construal: Overview, Motivation, and Some New Evidence. *Journal of Psycholinguistic Research, 26*(3), 277–295.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

Friederici, A. D. (2006). The Neural Basis of Language Development and Its Impairment. *Neuron, 52*(6), 941–952. https://doi.org/10.1016/j.neuron.2006.12.002

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, *22*(5), 1219–1248. https://doi.org/10.1037/0278-7393.22.5.1219

Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct Neurophysiological Patterns Reflecting Aspects of Syntactic Complexity and Syntactic Repair. *Journal of Psycholinguistic Research, 31*(1).

Friederici, A. D., Kotz, S. A., Werheid, K., Hein, G., & von Cramon, D. Y. (2003). Syntactic comprehension in Parkinson's disease: Investigating early automatic and late integrational processes using event-related brain potentials. *Neuropsychology*, *17*(1), 133–142. https://doi.org/10.1037/0894-4105.17.1.133

Friederici, A. D., Steinhauer, K., Mecklinger, A., & Meyer, M. (1998). Working memory constraints on syntactic ambiguity resolution as revealed by electrical brain responses. *Biological Psychology*, *47*(3), 193–221. https://doi.org/10.1016/S0301-0511(97)00033-1

Frisch, S., Schlesewsky, M., Saddy, D., & Alpermann, A. (2002). The P600 as an indicator of syntactic ambiguity. *Cognition*, *85*(3), B83–B92. https://doi.org/10.1016/S0010-0277(02)00126-9

Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K. J., Bastos, A. M., Pinotsis, D., & Litvak, V. (2015). LFP and oscillations—What do they tell us? *Current Opinion in Neurobiology*, *31*, 1–6. https://doi.org/10.1016/j.conb.2014.05.004

Gilbert, C. D., & Wiesel, T. N. (1983). Functional Organization of the Visual Cortex. In *Progress in Brain Research* (Vol. 58, pp. 209–218). Elsevier. https://doi.org/10.1016/S0079-6123(08)60022-9

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, *25*(2), 149–188. https://doi.org/10.1080/01690960902965951

Granena, G. (2013). Individual Differences in Sequence Learning Ability and Second Language Acquisition in Early Childhood and Adulthood. *Language Learning*, *63*(4), 665–703. https://doi.org/10.1111/lang.12018

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*(3), 311–343. https://doi.org/10.1177/0267658312461497

Gray, C. M., König, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, *338*(6213), 334–337. https://doi.org/10.1038/338334a0

Grey, S., Tanner, D., & van Hell, J. G. (2017). How right is left? Handedness modulates neural responses during morphosyntactic processing. *Brain Research*, *1669*, 27–43. https://doi.org/10.1016/j.brainres.2017.05.024

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (sps) as an erp measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439–483. https://doi.org/10.1080/01690969308407585

Hagoort, P., Wassenaar, M., & Brown, C. M. (2003). Syntax-related ERP-effects in Dutch. *Cognitive Brain Research*, *16*(1), 38–50. https://doi.org/10.1016/S0926-6410(02)00208-2

Hahne, A., Eckstein, K., & Friederici, A. D. (2004). Brain Signatures of Syntactic and Semantic Processes during Children's Language Development. *Journal of Cognitive Neuroscience, 16*(7), 1302–1318. https://doi.org/10.1162/0898929041920504

Hald, L. A., Bastiaansen, M. C. M., & Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language*, *96*(1), 90–105. https://doi.org/10.1016/j.bandl.2005.06.007

Harley, B., & Hart, D. (1997). Language Aptitude and Second Language Proficiency in Classroom Learners of Different Starting Ages. *Studies in Second Language Acquisition*, *19*(3), 379–400. JSTOR.

Herrmann, C. S., Munk, M. H. J., & Engel, A. K. (2004). Cognitive functions of gamma-band activity: Memory match and utilization. *Trends in Cognitive Sciences*, *8*(8), 347–355. https://doi.org/10.1016/j.tics.2004.06.006

Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, *19*(1), 59–73. https://doi.org/10.1016/j.cogbrainres.2003.10.022

Hohenstein, J., Eisenberg, A., & Naigles, L. (2006). Is he floating across or crossing afloat? Cross-influence of L1 and L2 in Spanish–English bilingual adults. *Bilingualism: Language and Cognition*, *9*(3), 249–261. https://doi.org/10.1017/S1366728906002616

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56. https://doi.org/10.1177/0267658312461803

Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, *32*(2), 277–307.

https://doi.org/10.1177/0267658315624960

Hwang, H., Shin, J.-A., & Hartsuiker, R. J. (2018). Late Bilinguals Share Syntax Unsparingly Between L1 and L2: Evidence From Crosslinguistically Similar and Different Constructions. *Language Learning*, *68*(1), 177–205. https://doi.org/10.1111/lang.12272

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149.

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, *4*(2), 257–282. https://doi.org/10.1075/lab.4.2.05kaa

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159–201. https://doi.org/10.1080/016909600386084

Kaan, E., & Swaab, T. Y. (2003). Repair, Revision, and Complexity in Syntactic Analysis: An Electrophysiological Differentiation. *Journal of Cognitive Neuroscience*, *15*(1), 98–110. https://doi.org/10.1162/089892903321107855

Kielar, A., Meltzer, J. A., Moreno, S., Alain, C., & Bialystok, E. (2014). Oscillatory Responses to Semantic and Syntactic Violations. *Journal of Cognitive Neuroscience*, *26*(12), 2840–2862. https://doi.org/10.1162/jocn_a_00670

Kielar, A., Meltzer-Asscher, A., & Thompson, C. K. (2012). Electrophysiological responses to argument structure violations in healthy adults and individuals with agrammatic aphasia. *Neuropsychologia*, *50*(14), 3320–3337. https://doi.org/10.1016/j.neuropsychologia.2012.09.013

Kielar, A., Shah-Basak, P. P., Deschamps, T., Jokel, R., & Meltzer, J. A. (2019). Slowing is slowing: Delayed neural responses to words are linked to abnormally slow resting state activity in primary progressive aphasia. *Neuropsychologia*, *129*, 331–347. https://doi.org/10.1016/j.neuropsychologia.2019.04.007

Kim, A., Oines, L., & Miyake, A. (2018). Individual differences in verbal working memory underlie a tradeoff between semantic and structural processing difficulty during language comprehension: An ERP investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(3), 406–420. https://doi.org/10.1037/xlm0000457

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205–225. https://doi.org/10.1016/j.jml.2004.10.002

Kim, A., & Sikos, L. (2011). Conflict and surrender during sentence processing: An ERP study of syntax-semantics interaction. *Brain and Language*, *118*(1–2), 15–22. https://doi.org/10.1016/j.bandl.2011.03.002

King, J., & Just, M. A. (1991). Individual Differences in Syntactic Processing: The Role of Working Memory. *Journal of Memory and Language; New York*, *30*(5), 580–602.

Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85*(1), 1–36. https://doi.org/10.1016/S0093-934X(02)00548-5

König, P., & Schillen, T. B. (1991). Stimulus-Dependent Assembly Formation of Oscillatory Responses: I. Synchronization. *Neural Computation*, *3*(2), 155–166. https://doi.org/10.1162/neco.1991.3.2.155

Kos, M., van den Brink, D., & Hagoort, P. (2012). Individual Variation in the Late Positive Complex to Semantic Anomalies. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00318

Kos, M., Vosse, T., van den Brink, D., & Hagoort, P. (2010). About Edible Restaurants: Conflicts between Syntax and Semantics as Revealed by ERPs. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00222

Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language*, *109*(2), 68–74. https://doi.org/10.1016/j.bandl.2008.06.002

Kroll, J. F., & Sunderman, G. (2003). Cognitive Processes in Second Language Learners and Bilinguals: The Development of Lexical and Conceptual Representations. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 104–129). Blackwell Pub.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49. https://doi.org/10.1016/j.brainres.2006.12.063

Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2020). A Tale of Two Positivities and the N400: Distinct Neural Signatures Are Evoked by Confirmed and Violated Predictions at Different Levels of Representation. *Journal of Cognitive Neuroscience*, *32*(1), 12–35. https://doi.org/10.1162/jocn_a_01465

Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. J. (2006). Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. *Language and Cognitive Processes*, *21*(5), 489–530. https://doi.org/10.1080/01690960500094279

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language*, *100*(3), 223–237. https://doi.org/10.1016/j.bandl.2005.12.006

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*(1), 117–129. https://doi.org/10.1016/S0926-6410(03)00086-7

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163. https://doi.org/10.1038/307161a0

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284. https://doi.org/10.1080/01638539809545028

Latent Semantic Analysis @ CU Boulder. (1998). LSA @ CU Boulder. http://lsa.colorado.edu/

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933. https://doi.org/10.1038/nrn2532

Lewis, A. G., & Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex*, *68*, 155–168. https://doi.org/10.1016/j.cortex.2015.02.014

Lewis, A. G., Schoffelen, J.-M., Schriefers, H., & Bastiaansen, M. (2016). A Predictive Coding Perspective on Beta Oscillations during Sentence-Level Language Comprehension. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00085

Li, X., Zhao, H., & Lu, Y. (2014). The Relation between Thematic Role Computing and Semantic Relatedness Processing during On-Line Sentence Comprehension. *PLoS ONE*, *9*(4). https://doi.org/10.1371/journal.pone.0095834

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00213

Luo, Y., Zhang, Y., Feng, X., & Zhou, X. (2010). Electroencephalogram oscillations differentiate semantic and prosodic processes during sentence reading. *Neuroscience*, *169*(2), 654–664. https://doi.org/10.1016/j.neuroscience.2010.05.032

Magyari, L., Bastiaansen, M. C. M., de Ruiter, J. P., & Levinson, S. C. (2014). Early Anticipation Lies behind the Speed of Response in Conversation. *Journal of Cognitive Neuroscience*, *26*(11), 2530–2539. https://doi.org/10.1162/jocn_a_00673

Makowski, D. (2018). The psycho Package: An Efficient and Publishing-Oriented Workflow for Psychological Science. *The Journal of Open Source Software*, *3*(22), 470. https://doi.org/10.21105/joss.00470

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., & Osterhout, L. (2010). Brain Potentials Reveal Discrete Stages of L2 Grammatical Learning. *Language Learning*, *60*(s2), 123–150. https://doi.org/10.1111/j.1467-9922.2010.00604.x

Meara, P. M. (2005). *Llama Language Aptitude Tests*. Lognostics.

http://www.lognostics.co.uk/tools/llama/

Mehravari, A. S., Emmorey, K., Prat, C. S., Klarman, L., & Osterhout, L. (2017). Brain-based individual difference measures of reading skill in deaf and hearing adults. *Neuropsychologia*, *101*, 153–168. https://doi.org/10.1016/j.neuropsychologia.2017.05.004

Mehravari, A. S., Tanner, D., Wampler, E. K., Valentine, G. D., & Osterhout, L. (2015). Effects of Grammaticality and Morphological Complexity on the P600 Event-Related Potential Component. *PLoS ONE*, *10*(10). https://doi.org/10.1371/journal.pone.0140850

Meulman, N., Stowe, L. A., Sprenger, S. A., Bresser, M., & Schmid, M. S. (2014). An ERP study on L2 syntax processing: When do learners fail? *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01072

Meyer, L., Obleser, J., & Friederici, A. D. (2013). Left parietal alpha enhancement during working memory-intensive sentence processing. *Cortex*, *49*(3), 711–721. https://doi.org/10.1016/j.cortex.2012.03.006

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2011). Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, *331*(6014), 176. https://doi.org/10.1126/science.1199644

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). Sequential Expectations: The Role of Prediction-Based Learning in Language. *Topics in Cognitive Science*, *2*(1), 138–153. https://doi.org/10.1111/j.1756-8765.2009.01072.x

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, *137*(2), 267–296.

https://doi.org/10.1037/a0021890

Moore, M., & Gordon, P. C. (2015). Reading ability and print exposure: Item response theory analysis of the author recognition test. *Behavior Research Methods*, *47*(4), 1095–1109. https://doi.org/10.3758/s13428-014-0534-3

Muñoz, M., & Cadierno, T. (2019). Mr Bean exits the garage driving or does he drive out of the garage? Bidirectional transfer in the expression of Path. *International Review of Applied Linguistics in Language Teaching*, *57*(1), 45–69. https://doi.org/10.1515/iral-2018-2006

Münte, T. F., Wieringa, B. M., Weyerts, H., Szentkuti, A., Matzke, M., & Johannes, S. (2001). Differences in brain potentials to open and closed class words: Class and frequency effects. *Neuropsychologia*, *39*(1), 91–102. https://doi.org/10.1016/S0028-3932(00)00095-6

Naiman, N. (Ed.). (1996). *The Good Language Learner* (Vol. 4). Multilingual Matters.

Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and Span: Working Memory Capacity Impacts the Use of Animacy but Not of World Knowledge during Spoken Sentence Comprehension. *Journal of Cognitive Neuroscience*, *22*(12), 2886–2898. https://doi.org/10.1162/jocn.2009.21400

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When Peanuts Fall in Love: N400 Evidence for the Power of Discourse. *Journal of Cognitive Neuroscience*, *18*(7), 1098–1111. https://doi.org/10.1162/jocn.2006.18.7.1098

Núñez-Peña, M. I., & Honrubia-Serrano, M. L. (2004). P600 related to rule violation in an arithmetic task. *Cognitive Brain Research*, *18*(2), 130–141. https://doi.org/10.1016/j.cogbrainres.2003.09.010

Nuwer, M. R., Comi, G., Emerson, R., Fuglsang-Frederiksen, A., Guérit, J.-M., Hinrichs, H., Ikeda, A., Jose C. Luccas, F., & Rappelsburger, P. (1998). IFCN standards for digital recording of clinical EEG. *Electroencephalography and Clinical Neurophysiology, 106*(3), 259–261. https://doi.org/10.1016/S0013-4694(97)00106-5

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP Study of Second Language Learning after Childhood: Effects of Proficiency. *Journal of Cognitive Neuroscience, 17*(8), 1212–1228. https://doi.org/10.1162/0898929055002436

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). *FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data* [Research Article]. Computational Intelligence and Neuroscience; Hindawi. https://doi.org/10.1155/2011/156869

O'Rourke, P., & Colflesh, G. (2015). P600 Dominance Predicts Comprehension of Garden-Path Sentences. *LSA Annual Meeting Extended Abstracts, 6*(0), 17-1–5. https://doi.org/10.3765/exabs.v0i0.3008

Osterhout, L. (1997). On the Brain Response to Syntactic Anomalies: Manipulations of Word Position and Word Class Reveal Individual Differences. *Brain and Language, 59*(3), 494–522. https://doi.org/10.1006/brln.1997.1793

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*(6), 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z

Osterhout, L., & Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes, 8*(4), 413–437. https://doi.org/10.1080/01690969308407584

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain Potentials Elicited by Garden-

Path Sentences: Evidence of the Application of Verb Information During Parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 786–803.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice Learners, Longitudinal Designs, and Event-Related Potentials: A Means for Exploring the Neurocognition of Second Language Processing. *Language Learning*, *56*(s1), 199–230. https://doi.org/10.1111/j.1467-9922.2006.00361.x

Osterhout, L., & Mobley, L. A. (1995). Event-Related Brain Potentials Elicited by Failure to Agree. *Journal of Memory and Language*, *34*, 739–773.

Osterhout, L., & Nicol, J. (1999). On the Distinctiveness, Independence, and Time Course of the Brain Responses to Syntactic and Semantic Anomalies. *Language and Cognitive Processes*, *14*(3), 283–317. https://doi.org/10.1080/016909699386310

Osterhout, L., Pitkänen, I., McLaughin, J., & Zeitlin, M. (2019). Event-related Potentials as Metrics of Foreign Language Learning and Loss. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition* (pp. 402–416). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.32

Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., Frenck-Mestre, C., & Hirschensohn, J. (2008). Second-language learning and changes in the brain. *Journal of Neurolinguistics*, *21*(6), 509–521. https://doi.org/10.1016/j.jneuroling.2008.01.001

Pakulak, E., & Neville, H. J. (2010). Proficiency Differences in Syntactic Processing of Monolingual Native Speakers Indexed by Event-related Potentials. *Journal of Cognitive Neuroscience*, *22*(12), 2728–2744. https://doi.org/10.1162/jocn.2009.21393

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681.

https://doi.org/10.1038/nn1082

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing Syntactic Relations in Language and Music: An Event-Related Potential Study. *Journal of Cognitive Neuroscience*, *10*(6), 717–733. https://doi.org/10.1162/089892998563121

Pefkou, M., Arnal, L. H., Fontolan, L., & Giraud, A.-L. (2017). θ-Band and β-Band Neural Activity Reflects Independent Syllable Tracking and Comprehension of Time-Compressed Speech. *The Journal of Neuroscience*, *37*(33), 7930–7938. https://doi.org/10.1523/JNEUROSCI.2882-16.2017

Penolazzi, B., Angrilli, A., & Job, R. (2009). Gamma EEG activity induced by semantic violation during sentence reading. *Neuroscience Letters*, *465*(1), 74–78. https://doi.org/10.1016/j.neulet.2009.08.065

Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology*, *15*(2), 202–206. https://doi.org/10.1016/j.conb.2005.03.007

Pérez, A., Molinaro, N., Mancini, S., Barraza, P., & Carreiras, M. (2012). Oscillatory dynamics related to the Unagreement pattern in Spanish. *Neuropsychologia*, *50*(11), 2584–2597. https://doi.org/10.1016/j.neuropsychologia.2012.07.009

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*(3), 407–428. https://doi.org/10.1016/j.cogbrainres.2004.09.012

Pijnacker, J., Geurts, B., van Lambalgen, M., Buitelaar, J., & Hagoort, P. (2010). Exceptions and anomalies: An ERP study on context sensitivity in autism. *Neuropsychologia*, *48*(10), 2940–2951. https://doi.org/10.1016/j.neuropsychologia.2010.06.003

Pimsleur, P., Reed, D., & Stansfield, C. W. (2004). *Pimsleur Language Aptitude Battery*. Language Learning and Testing Foundation.

Predictors of Postsecondary Success. (2013). American Institutes for Research.

Prystauka, Y., & Lewis, A. G. (2019). The power of neural oscillations to inform sentence comprehension: A linguistic perspective. *Language and Linguistics Compass, 13*(9), e12347. https://doi.org/10.1111/lnc3.12347

Purdy, J. D., Leonard, L. B., Weber-Fox, C., & Kaganovich, N. (2014). Decreased Sensitivity to Long-Distance Dependencies in Children With a History of Specific Language Impairment: Electrophysiological Evidence. *Journal of Speech, Language, and Hearing Research*, *57*(3), 1040–1059. https://doi.org/10.1044/2014_JSLHR-L-13-0176

Qi, Z., Beach, S. D., Finn, A. S., Minas, J., Goetz, C., Chan, B., & Gabrieli, J. D. E. (2017). Native-language N400 and P600 predict dissociable language-learning abilities in adults. *Neuropsychologia*, *98*, 177–191. https://doi.org/10.1016/j.neuropsychologia.2016.10.005

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2016). N400 amplitudes reflect change in a probabilistic representation of meaning: *Proceedings of the Annual Meeting of Cognitive Science Society*, 2045–2050.

Robinson, P. (2002). *Individual Differences and Instructed Language Learning*. John Benjamins Publishing.

Rogers, V. E., Meara, P., Aspinall, R., Fallon, L., Goss, T., Keey, E., & Thomas, R. (2016). Testing aptitude: Investigating Meara's (2005) LLAMA tests. *EUROSLA Yearbook*, *16*, 179–210. https://doi.org/10.1075/eurosla.16.07rog

Rogers, V. E., Meara, P., Barnett-Legh, T., Curry, C., & Davie, E. (2017). Examining the LLAMA aptitude tests. *Journal of the European Second Language Association*, *1*(1), 49–60. https://doi.org/10.22599/jesla.24

Rommers, J., Dijkstra, T., & Bastiaansen, M. (2012). Context-dependent Semantic Processing in the Human Brain: Evidence from Idiom Comprehension. *Journal of Cognitive Neuroscience*, *25*(5), 762–776. https://doi.org/10.1162/jocn_a_00337

Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006a). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience,* 2030–2048.

Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006b). The Impact of Proficiency on Syntactic Second-language Processing of German and Italian: Evidence from Event-related Potentials. *Journal of Cognitive Neuroscience*, *18*(12), 2030–2048. https://doi.org/10.1162/jocn.2006.18.12.2030

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high-and low-frequency words. *Memory & Cognition*, *18*(4), 367–379. https://doi.org/10.3758/BF03197126

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research*, *24*(3), 397–430. https://doi.org/10.1177/0267658308090186

Schneider, J. M., & Maguire, M. J. (2018). Identifying the relationship between oscillatory dynamics and event-related responses. *International Journal of Psychophysiology*, *133,* 182–192. https://doi.org/10.1016/j.ijpsycho.2018.07.002

Segaert, K., Mazaheri, A., & Hagoort, P. (2018). Binding language: Structuring sentences through precisely timed oscillatory mechanisms. *European Journal of Neuroscience, 48*(7), 2651–2662. https://doi.org/10.1111/ejn.13816

Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., & Kuperberg, G. R. (2008). Two Neurocognitive Mechanisms of Semantic Integration during the Comprehension of Visual Real-world Events. *Journal of Cognitive Neuroscience*, *20*(11), 2037–2057. https://doi.org/10.1162/jocn.2008.20143

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Skehan, P., & Ducroquet, L. (1988). *A comparison of first and foreign language ability* [Working Documents No. 8]. ESOL Department, Institude of Education, London University.

Sparks, R. L., Humbach, N., & Javorsky, J. (2008). Individual and longitudinal differences among high and low-achieving, LD, and ADHD L2 learners. *Learning and Individual Differences*, *18*(1), 29–43. https://doi.org/10.1016/j.lindif.2007.07.003

Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2009). Long-Term Crosslinguistic Transfer of Skills From L1 to L2. *Language Learning*, *59*(1), 203–243. https://doi.org/10.1111/j.1467-9922.2009.00504.x

Stanovich, K. E., & West, R. F. (1989). Exposure to Print and Orthographic Processing. *Reading Research Quarterly*, *24*(4), 402–433. JSTOR. https://doi.org/10.2307/747605

Steinhauer, K. (2014). Event-related Potentials (ERPs) in Second Language Research: A Brief Introduction to the Technique, a Selected Review, and an Invitation to Reconsider Critical Periods in L2. *Applied Linguistics*, *35*(4), 393–417. https://doi.org/10.1093/applin/amu028

Sulpizio, S., Del Maschio, N., Fedeli, D., & Abutalebi, J. (2020). Bilingual language processing:

A meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *108*, 834–853. https://doi.org/10.1016/j.neubiorev.2019.12.014

Sunderman, G., & Kroll, J. F. (2006). First Language Activation During Second Langauge Lexical Processing: An Investigation of Lexical Form, Meaning, and Grammatical Class. *Studies in Second Language Acquisition*, *28*(3), 387–422. JSTOR.

Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement: A Commentary on “Grammatical agreement processing in reading: ERP findings and future directions” by Molinaro et al., 2014. *Cortex*, *66*, 149–155. https://doi.org/10.1016/j.cortex.2014.04.007

Tanner, D. (2018). *General files for “Robust neurocognitive individual differences in grammatical agreement processing: A latent variable approach”* (Version 2) [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/DKEKBH

Tanner, D. (2019). Robust neurocognitive individual differences in grammatical agreement processing: A latent variable approach. *Cortex*, *111*, 210–237. https://doi.org/10.1016/j.cortex.2018.10.011

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition*, *17*(2), 277–293. https://doi.org/10.1017/S1366728913000370

Tanner, D., Mclaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, *16*(2), 367–382. https://doi.org/10.1017/S1366728912000302

Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, *56*, 289–301.

https://doi.org/10.1016/j.neuropsychologia.2014.02.002

The State of Languages in the U.S.: A Statistical Portrait. (2016). American Academy of Arts and Sciences.

Ulbrich, C., & Ordin, M. (2014). Can L2-English influence L1-German? The case of post-vocalic /r/. *Journal of Phonetics*, *45*, 26–42. https://doi.org/10.1016/j.wocn.2014.02.008

van de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., & Vissers, C. T. W. M. (2009). Monitoring in Language Perception. *Language and Linguistics Compass*, *3*(5), 1211–1224. https://doi.org/10.1111/j.1749-818X.2009.00163.x

van de Meerendonk, N., Kolk, H. H. J., Vissers, C. Th. W. M., & Chwilla, D. J. (2010). Monitoring in Language Perception: Mild and Strong Conflicts Elicit Different ERP Patterns. *Journal of Cognitive Neuroscience*, *22*(1), 67–82. https://doi.org/10.1162/jocn.2008.21170

van Hell, J. G., & Tokowicz, N. (2010). Event-related brain potentials and second language learning: Syntactic processing in late L2 learners at different L2 proficiency levels. *Second Language Research*, *26*(1), 43–74. https://doi.org/10.1177/0267658309337637

van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, *22*(2), 241–255. https://doi.org/10.1016/j.cogbrainres.2004.09.002

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequencyinevent-related brainpotentials. *Memory & Cognition*, *18*(4), 380–393. https://doi.org/10.3758/BF03197127

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190.

https://doi.org/10.1016/j.ijpsycho.2011.09.015

Venables, W. N., & Riple, B. D. (2002). *Modern applied statistics with S*. Springer.

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., & Cappa, S. F. (2011). Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience & Biobehavioral Reviews*, *35*(3), 407–426. https://doi.org/10.1016/j.neubiorev.2010.04.007

Vignali, L., Himmelstoss, N. A., Hawelka, S., Richlan, F., & Hutzler, F. (2016). Oscillatory Brain Dynamics during Sentence Reading: A Fixation-Related Spectral Perturbation Analysis. *Frontiers in Human Neuroscience, 10*. https://doi.org/10.3389/fnhum.2016.00191

Wang, L., Jensen, O., Brink, D. van den, Weder, N., Schoffelen, J.-M., Magyari, L., Hagoort, P., & Bastiaansen, M. (2012). Beta oscillations relate to the N400m during language comprehension. *Human Brain Mapping*, *33*(12), 2898–2912. https://doi.org/10.1002/hbm.21410

Wang, L., Zhu, Z., & Bastiaansen, M. (2012). Integration or Predictability? A Further Specification of the Functional Role of Gamma Oscillations in Language Comprehension. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00187

Wang, X.-J. (2010). Neurophysiological and Computational Principles of Cortical Rhythms in Cognition. *Physiological Reviews*, *90*(3), 1195–1268. https://doi.org/10.1152/physrev.00035.2008

Wassenaar, M., Brown, C. M., & Hagoort, P. (2004). ERP Effects of Subject—Verb Agreement Violations in Patients with Broca’s Aphasia. *Journal of Cognitive Neuroscience*, *16*(4), 553–576. https://doi.org/10.1162/089892904323057290

Weber-Fox, C., Davis, L. J., & Cuadrado, E. (2003). Event-related brain potential markers of high-language proficiency in adults. *Brain and Language*, *85*(2), 231–244. https://doi.org/10.1016/S0093-934X(02)00587-4

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational Constraints on Functional Specializations for Language Processing: ERP and Behavioral Evidence in Bilingual Speakers. *Journal of Cognitive Neuroscience*, *8*(3), 231–256. https://doi.org/10.1162/jocn.1996.8.3.231

Weiss, S., & Mueller, H. M. (2003). The contribution of EEG coherence to the investigation of language. *Brain and Language*, *85*(2), 325–343. https://doi.org/10.1016/S0093-934X(03)00067-1

Weiss, S., Mueller, H. M., Schack, B., King, J. W., Kutas, M., & Rappelsberger, P. (2005). Increased neuronal communication accompanying sentence comprehension. *International Journal of Psychophysiology*, *57*(2), 129–141. https://doi.org/10.1016/j.ijpsycho.2005.03.013

Wells, G. (1985). *Language Development in the Pre-School Years*. Cambridge University Press.

West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research*, *13*(3), 363–375. https://doi.org/10.1016/S0926-6410(01)00129-X

White, E. J., Genesee, F., & Steinhauer, K. (2012). Brain Responses before and after Intensive Second Language Learning: Proficiency Based Changes and First Language Background Effects in Adult Learners. *PLoS ONE*, *7*(12). https://doi.org/10.1371/journal.pone.0052318